Sensitivity, Uncertainty and Refinement of a Global Flood Model

Cain Thomas Moylan

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Physical Geography (PhD) in the Faculty of Science

Word Count: 65,166

Abstract

Floods are a major natural hazard, causing significant damage and fatalities across the world. The modelling of floods is possible, but has been historically performed by developed nations, with good access to data, allowing them to build and test hydraulic models which can predict the frequency and severity of flooding events. However, in recent years, satellite applications have led to a proliferation of remotely sensed data products, hence making it feasible to run global scale flood models. However, while feasible, the new generation of global flood models face serious issues, as the datasets are coarse and there are significant assumptions required for consistent methodologies. It is therefore necessary to understand the effect that these assumptions have on flood hazard predictions. In this thesis, one of the new generation of global flood models is analysed. First, the uncertainty of the model's parameters is assessed, which codify assumptions into the modelling methodology. From a pool of 36 parameters, it is found the vast majority of the model's output variance can be represented with only 7 parameters. Following this, the model is tested more rigorously at a single case-study location, the Po river basin in north Italy. Here the model is assessed for its ability to represent local conditions, and it is found that even a global scale model requires some information at the local scale to appropriately constrain the uncertainty of the predictions. Once done, there are modelling outputs of reasonable skill. The final chapter focusses on the prediction of peak flows, found to be the most uncertain component in the methodology. The regionalization scheme of the model was refined to incorporate uncertainty estimation. This thesis has therefore started the long process of incorporating uncertainty into global flood modelling and serves as a guide for further work in the field.

Acknowledgements

This volume would have been impossible to achieve on my own, without the guidance and support of many people.

To my supervisors Jeff and Jim, I offer my thanks for introducing me to such an interesting area of study and providing me with the resources I needed to get the job done, while offering rigorous commentary throughout.

Thanks to Thorsten Wagener, Ross Woods and the rest of the WISE CDT staff for accepting me onto the course, I have had the opportunity to learn a huge amount and has opened my world to hitherto unknown possibilities. I am also deeply grateful to Alberto Montanari and his colleagues at DICAM for making me feel so welcome during my placement at Università di Bologna. It was a steep learning experience but one I will never forget. Grazie di tutto a voi!

This PhD would have been much greyer without the discussions and banter which I had with my peers on this course (French squad I am looking at you) and sharing a house with Ludo and Anna certainly kept me on the straighter path at times! My office hours were also made a lot less gruelling by Simbi's presence as a desk-mate. Regular tea and ping pong breaks made the time fly by (for better or worse).

Special thanks to Laurie for immediately agreeing to proofread the manuscript for me. It has certainly avoided any future readers having some rather confusing moments. Another thanks to him and Rory for being great housemates and providing excellent distractions at the times it mattered most.

There were some extremely difficult moments throughout this process and I was able to maintain a greater stability throughout it all thanks to the support and guidance of Isabel de Salis.

My parents also deserve thanks always encouraging me to push for the highest level of achievement I can reach, giving me the confidence to define my own goals and follow them. Thanks to Mum and John as well for supporting me when the stipend ran out and letting me come back for a few weeks when my rent expired, providing some absolutely essential stability and support.

A special thankyou as well to Elisa, for everything which she has shared and helped me with over the past 3 years. They have been invaluable moments.

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: CAIN MOYLAN DATE: 01 / 08 / 2020

Table of Contents

List of Figures

List of Tables

1. Introduction

"A little wave came rolling past. He cut at it with his sword and laughed contemptuously. Then he coolly retired and ordered the "general engagement" to be sounded. The archers shot, the slingers slung, the javelin-men threw their javelins; the infantry waded into the waters as far as their arm-pits and hacked at the little waves, the cavalry charged on either flank and swam out some way, slashing with their sabres, the mangonels hurled rocks and the catapults huge javelins and iron-tipped beams. Caligula then put out to sea in a war-vessel and anchored just out of range of the missiles, uttering absurd challenges to Neptune and spitting far out over the vessel's side."

Caligula declares war on Neptune – Robert Graves, *I, Claudius*

Flooding has had a profound impact on human life throughout the course of history. The mythological motif of the flood is pervasive across ancient cultures, with over 150 unique examples provided by Isaak (2002). Although in modern terms they are often conceived of as calamitous events, in older stories, alongside the destruction that great floods brought a sense for renewal and the creation of the world, as the power of water became deified (Isaak, 2002). Indeed, it is simply the case that a complete anthropological account of humanity is not possible without a discussion of our relationship to water and flooding. While the mythos of flooding has seeped out of modern discussion, there is no question that it continues to represent a significant cultural and societal issue. As our cities have continued to develop and expand close to water (UNISDR, 2015), we find even the most well-developed of them can become inundated at great expense to the inhabitants of the city and the governing institutions (Hinkel et al., 2014). However, rather than modern discourses about flooding being conducted in mythological terms, their significance is instead recognised through the terminology of casualties, cost and risk.

When thought of in these terms, the problems of flooding become much more severe. Every year, floods cause thousands of fatalities, affect millions, and cost billions (EM-DAT, 2018), and up to a third of all recorded natural hazards around the globe are flood related. However, despite the ominous threat which flooding represents, there are still reasons to be optimistic. As our conceptions of risk have developed, so have our conceptions of mitigation (UNISDR,

2015). As we have learnt more about the underlying mechanisms which cause flooding, we have been able to effectively model these mechanisms via hazard modelling and flood forecasting (Alfieri et al., 2013; Schumann et al., 2014). This has reached a degree of sophistication such that many developed countries have national scale flood warning systems, such as the Met Office (Williams et al., 2018) for the UK, or Flood Early Warning Systems for the Netherlands (Berendrecht et al., 2011; Gijsbers et al., 2017). Indeed, there is now even a flood forecasting system which provides forecasts for the entire planet, the Global Flood Awareness System (GloFAS) (Emerton et al., 2018).

There are several research groups which are working on models which attempt to quantify flood hazard and flood risk at the global scale. Such models are known as global flood models (GFMs), and they have a great potential to aid in the mitigation of flood risk around the globe (Ward et al., 2015). Such capabilities have recently become possible due to the proliferation of remotely sensed data sets needed by modellers to predict extreme events across the globe (Di Baldassarre & Uhlenbrook, 2012). These data sets also allow the use of globally consistent data, which potentially reduces the great geographical disparities in understanding of flood risk and the consequences that it brings (Tanoue et al., 2016). These developments surely indicate the beginning of global-scale inundation modelling, as surely as it was once said for hydrology (Eagleson, 1986).

While the development of global flood modelling as a discipline should certainly be treated as a success (Schumann et al., 2018), the community is still yet to establish any clear best practice or deep understanding of model uncertainty. On the contrary, it has been found that there are fundamental differences between the modelling approaches used, and they do not produce similar estimates of hazard (Bernhofen et al., 2018; Trigg et al., 2016). It has been noted that a framework for validation is needed if the global flood modelling field is to continue its development, which puts an emphasis on understanding the differences between models (Hoch & Trigg, 2019) and while it has been shown in isolated cases that the uncertainties of the GFMs can be identified (Eilander et al., 2018; Grimaldi et al., 2019), they are yet to be systematically addressed. It is for this purpose that the following research is conducted and aims to contribute towards a more rigorous understanding of global flood modelling.

1.1. Aims and objectives

The overall aim of the thesis is to provide a contribution to the field of global flood modelling, focussing on the parameter sensitivities and uncertainties of a single GFM. While recent inter-model comparisons have highlighted that there are disparities in the models and therefore our understanding of the systems being modelled, we are unable to say which models are more accurate and why without first understanding the behaviour of the individual GFMs. Therefore, the research of this thesis will be within the remit of a single global flood model: the model developed at the University of Bristol, henceforth to be referred to as the Bristol GFM, although in current literature it is also sometimes referred to as the SSBN model (Sampson et al., 2015) or Fathom-Global flood model. Despite the focus on this single model, many of the other GFMs face the same limitations in terms of computational resources, use of remotely sensed data and a limited suite of methods for flood peak prediction. Thus, it is anticipated that the conclusions reached herein will have implications for the other models within the discipline.

Through the investigations of the Bristol GFM, there have emerged some key research questions. Initial investigations within this thesis were very open ended and centred on understanding the sensitivity of model outputs to all model parameters. However, subsequent studies were then centred around more precise aspects of the Bristol GFM focusing on the most sensitive parameters given these initial analyses.

Research questions:

1: What are the fundamental drivers of variability in output within the Bristol GFM?

In the first investigation, the primary interest was simply to understand better what exactly the model was doing and how much influence model parameters have on flood hazard. Prior to this study, a full list and descriptions of the parameters did not yet exist, and neither was it known to what extent the parameters would influence model outputs. Hence this meant it was necessary to map out the structure of the model and create an inventory of the model parameters, presented in chapter 3. With this information gathered, a sensitivity analysis was conducted on the model, to understand what effect the numerous parameters within the model have on model output. To this end, the analysis was conducted at various locations to account for the effects of location on input data and parameter sensitivity, and the variance

of model outputs was used to assess which parameters could be classed as important and which could not. A full description of the sensitivity analysis technique used can be found in chapter 4, while the application and results are in chapter 5. This task itself required a thorough mapping of the model structure and an inventory of model parameters required for functioning. This is presented in chapter 3, in advance of the main data and methods section.

2: How well can a global-scale model represent a detailed local study?

Following on from the initial investigation, it was found that it is indeed possible to represent the vast majority of the model's output variance using a much smaller pool of model parameters. This allowed us to investigate in greater detail the behaviour of these parameters, to fully understand what effect they have on model output rather than simply understanding the lumped effect. In chapter 6, the model was evaluated in a Monte Carlo sampling methodology, with its performance evaluated at the Po river. We were interested in the extent to which the global model might be able to recreate a local scale model and observation data. It is expected that the global model's performance will lag behind the local model, meaning we are also interested to know what local information would be necessary to leverage a better performance. Within this research, the question of parameter uncertainty and the effect this has on model predictions again arises. A corresponding question is whether the optimisation of the model parameters to local performance metrics, i.e. correct delineation of floodplain extent, comes at the expense of the model not realistically representing other aspects of the system (e.g. water surface elevations).

3: How can we better quantify the uncertainty of flood peak estimation and improve prediction?

By examining the model in greater detail at a local study, it was found that the principle source of uncertainty and error within the model comes from the estimation of peak flow. Therefore, the chapter 7 focusses on assessing the extent of this uncertainty and whether it could be more appropriately constrained. In the first part of this chapter, the uncertainty of the peak flow prediction method was estimated, something which has not yet been done for global inundation modelling. Following on from this, several changes to the peak flow prediction method were implemented, with the aim of both reducing the uncertainty of the flow estimates and improving the actual estimates themselves. As the peak flow method is a flood

frequency analysis technique, it relies upon topographical and climatological attributes to pool gauged basins and make predictions of peak flows. The proposed alternative considered an alternative scheme of climate classification, as climate was flagged as a principal source of uncertainty and of interest to investigate. The flow prediction method will then be assessed with the old and new climate schemes, both in terms of predictive skill and overall uncertainty estimation.

1.2. Structure of thesis

The aim of this thesis is to assess elements of parameter sensitivity and uncertainty within a global flood model and then go through a stage of parameter refinement guided by an increased understanding of which parameters influence model outputs. To reach this point, a review of the current literature on inundation modelling and methods for modelling sensitivity is presented, including discussions on uncertainty, the necessary components of a flood inundation hazard model, including flood frequency analysis. This is followed in chapter 3 by a detailed description of the various functions within the Bristol GFM code. This is a standalone chapter, as a detailed technical understanding of the model will be necessary to understand the significance of the results chapters. After this, chapter 4 contains a description of the datasets used throughout the thesis and a brief description of methods. These methods are explained in fuller detail in the results chapters 5-7, as their precise implementation is dependent on the aims of the chapter. Each of the results chapters is themed around sensitivity, uncertainty and parameter refinement respectively, which build upon the work of the previous chapter. Following this is a discussion of the implications of the work, and to what extent the initial aims have been satisfied. The thesis closes with a synthesis of the overall contribution to the field of global flood modelling, and recommendations for further research.

2. Literature Review

In the following sections, the aim is to provide an overview of the various practical and theoretical foundations which are required for inundation modelling at the global scale. It also outlines some of the trends in the community of global flood modelling, and exposes the linkages with the implications for society at large. Brief overviews of the important theory for various methods have also been given. Certain sections, such as that about flood frequency analysis, will be covered in greater detail as they become relevant in later chapters.

2.1. Introduction to flooding in a global context

Flooding is a phenomenon which happens globally and has happened consistently throughout the length of human history. Into the 20th century, it is still possible to find examples of flooding in which tens of thousands of lives have been lost (Sandesara & Wooten, 1986). Indeed, flooding represents 31% of all recorded natural disasters, consistently representing the largest number of events worldwide (UNISDR, 2015). Furthermore, in 2018 of 315 recorded disasters, 127 of them were floods. These floods affected 34 million people worldwide, causing an estimated 2,879 fatalities and $19.7 billion dollars in damages (EM-DAT, 2018). The effects of flooding are also expected to become worse, with climate change driving more extreme events in some cases (Alfieri et al., 2015), while the effects of continued socio-economic growth and development is also a substantial driver of potential future risks, possibly leading to a 20-fold increase in current losses by 2100 (Winsemius et al., 2016) - a prediction shared for all natural disasters (Dilley, et al., 2005). Further to this, many of these growing population centres are characterised by informally planned settlements, compounding the effect of losses (Di Baldassarre et al., 2010). Clearly, this is an area which requires concerted focus on the part of decision makers.

With such a high frequency of disasters, the diversity of flood generating mechanisms and potential loss of value to society must be recognised – for example, there is substantial spatiotemporal variation in flood vulnerability across the globe (Tanoue et al., 2016). Within the hydrological community, the Panta Rhei decade of research is centred around human interactions with the natural environment, including the impact of human activity on the severity of flooding (Montanari et al., 2013). Meanwhile, the emergence of a new hydrological sub-field, that of global flood modelling (Schumann et al., 2018; Ward et al., 2015), testifies

to the crucial role that accurate quantification of flood risk plays in contemporary society, and there are regular calls within the scientific community which suggest as much (Schumann et al., 2014; Vorogushyn et al., 2018).

2.1.1. Concept of flooding and drivers of flooding mechanisms

Flooding is a temporary natural process, characterized by an excess of water over a landmass, which would normally not be inundated. There are many examples where flooding is a beneficial activity, such the lower Mississippi basin, where annual inundations deliver millions of tons of sediment, ensuring that the Louisiana coastline is naturally bolstered against the effects of the encroaching tides (Julien & Vensel, 2005). In such cases, the effects of levee systems hinder natural processes which have their own dynamic feedback, meaning other potential positives of flooding, such as its importance for soil health, is missed (Ponnamperuma, 1984). Often the only interaction which we collectively have with the activity of flooding is through the news, and it is typified as disastrous. This is the reason for viewing the phenomenon of flooding through the lens of risk.

The processes which lead to flooding can often be unpredictable and complex. The exact particulars of a location, along with specific initial conditions, leads to the problem of uniqueness of place for a hydrological modeller, often confounding a simple appraisal of events (Beven, 2000). It has been noted that the mechanisms which underlie flooding, while all being recognised in individual instances, fail to be connected in a singular conceptual framework, which impedes the effectiveness of our flood mitigation and adaptive strategies (Tarasova et al., 2019). Indeed, this is symptomatic of a systemic issue in hydrology pertaining to a lack of standardised classification (Wagener et al., 2007).

Within the context of an individual country, the range of flood generating mechanisms can be quite large. Even within single mountainous catchments, there can be ambiguity about the drivers of flooding (Sikorska et al., 2015). Within the United States, the effect of regional behaviour and weather conditions leads to a variety of dominating flood controls such as precipitation intensity and percentage of precipitation as snow, with a key indicator being soil moisture (Berghuijs et al., 2016). These differences also affect the seasonality and magnitude of events across the states, and the risk factor is exacerbated by the effects of urbanisation (Villarini, 2016).

The effects of such controls can be witnessed in a more global sense with reference to the contrasting events of the 2019 Iran floods with the 2011 Thailand floods. Within a particularly arid region, previously susceptible to drought for the last 30 years, several regions of Iran saw extreme events happen in the form of high intensity rainfall, caused by meteorological patterns undergoing changes due to climate change (Sharifi & Bokaie, 2019). With some areas receiving over 300mm of rain within the period of a single event, the intensity of the rainfall led to flooding generated by infiltration excess; typical for arid regions (Huang et al., 2016). This event caused over $2.2 billion of damage and had 76 fatalities (Al-Jazeera, 2019). By contrast, the Thailand flooding event was not atypical for the region in the sense of any unexpected processes. Rather, there was a particularly heavy series of storm events during the monsoon season, combined with the decadal occurrence of El Nino, leading to a series of floods around Bangkok, best characterised by saturation excess (Gale & Saunders, 2013). The consequence of this event was that it affected 13 million people mainly in the lower Chao Phraya basin, and caused in excess of $50 billion of damage, raising many questions about the resilience of the Thai infrastructure and whether the impact of such an event could have been preventable (World Bank, 2012).

2.1.2. Flood hazard and disaster risk

When thinking about the impact that flooding has on us, the most widespread understanding of flooding comes in the form of hazard. That is, the probabilistic measure of how likely and area is to flood, often having a damaging effect (Dankers et al., 2014). However, to characterise risk means accounting for the effect of the hazard at a given location. An inundation of the same volume is not the same risk, if in one scenario a rural area is flooded and in another it is an urban area. To account for this, the concepts of vulnerability and exposure are needed (Tanoue et al., 2016). From these, the concept of risk can be represented by the classic equation:

$$Risk = Hazard \times Exposure \times Vulnerability$$

Where the lumped metric of our concern for flooding is represented by risk. The hazard is the probability of flooding, while exposure is usually the value assets or number of people that might be affected by the hazard. Alongside exposure, vulnerability is a measure of the susceptibility of people or assets to harm or damage (UNISDR, 2015). This complicates the estimation of risk with respect to flooding, given that there are multiple interacting variables

which when combined determines the overall risk. For instance, hazard and exposure are considered areas of increasing concern, with extreme events expected to double (L. Alfieri et al., 2015), and the growth of informal settlements and socioeconomic developments occurring (Jongman et al., 2014; Winsemius et al., 2016), yet at the same time, the implementation of adaptation measures is resulting a broad reduction of vulnerability when effective (Haer et al., 2019; Jongman et al., 2015).

Within the case of flood modelling, the general modelling method is to look specifically at the hazard associated with floods, and the exposure and vulnerability are normally represented by other models (Teng et al., 2017). A major example of this is the Aqueduct flood risk analyser (Winsemius et al., 2015), which represents the total losses as well associated with a given return period in a country, given the hazard. But this is itself driven by the predictions produced by a global flood hazard model (Winsemius et al., 2013).

2.1.3. Sendai Framework

From 2015-2030, the UN is directing a global initiative to have all states and related shareholders direct their attention at the question of disaster risk reduction, through improved governance, management and resilience of the various constituents of natural disasters. This initiative is the Sendai Framework. One of the guiding principles of the framework is:

'Addressing underlying risk factors cost-effectively through investment versus relying primarily on post-disaster response and recovery'

It is from this position that we are looking to underline the importance of the following work. As mentioned, within the context of flood modelling, the major preoccupation is hazard. Hence with improved modelling of hazard, there will be an improved conception of flood risk across the globe. Recently, it was found that there was a vast underestimation of the flood risk of the United States, given the ad hoc distribution of flood studies conducted by FEMA across the country (Wing et al., 2018). By the addition of hazard maps in areas which have not before been modelled, it has been claimed that there are up to three times as many Americans at risk of flooding as when local methods have been used. Should such a technique be represented globally with newer alternative population data sets (Smith et al., 2019), it would represent a step-change and offer a complete reappraisal of what the global response

to flooding should be (Vorogushyn et al., 2018), providing an active platform through which to assess risk to flooding, rather than being reactive to risk as it happens.

2.1.4. The 'credibility challenge'

Despite the successes of continental scale modelling in the United States, there are major barriers to repeating this on the global scale, as the data quality begins to diminish (Crochemore et al., 2019). Further to this, the success of any modelling in the United States might not be readily transferable to the rest of the world (Hundecha et al., 2016). Indeed, the best representation of global flood hazard, while showing some successes, are still being held back by deep uncertainties and a general lack of process representation or complexity which compliments the natural behaviour of floods (Ward et al., 2015).

The significant obstacles faced by the global flood modelling community has best been demonstrated by a continental scale comparison of six of the major GFMs across the entirety of Africa (Trigg et al., 2016). What was found is that the models generally do not agree on the flood hazard, except in the most predictable circumstances where the floodplain is constrained by steep topography. More worrying is that in critical areas of Africa, such as the Nile and Niger deltas, there were extremely poor agreements between the models. In another study assessing the validity of the flood models, Bernhofen et al. (2018) found that across hydraulically diverse catchments there was a corresponding diversity of model performance. Furthermore, it was found that the most consistent way to achieve acceptable performance was to produce an ensemble of predictions from the best individual models, thus utilising the most appropriate aspects of each model in each circumstance.

The implications for these findings are that more systemic assessment of each GFM and detailed comparisons between each GFM and their respective predictions are needed (Hoch & Trigg, 2019). Although Hoch and Trigg are right to highlight the need for a GFM Validation Framework, their main preoccupation is with systematising comparisons between the models in a deterministic manner. In their proposal, there is not enough consideration given to the nature of model uncertainty or identification (Wagener & Gupta, 2005). With this in mind, the ideal next step would be for the community to augment their framework with some of the principles laid out by the broader hydrological Framework for the Understanding of Structural Errors (Clark et al., 2008) In this study, the various modelling assumptions of different hydrological models were considered as modules, to allow for the direct comparison of

process, rather than lumped comparisons of results. By following the same approach, the various modular components of the GFMs, of which there are many, could be directly compared and integrated, such that they are most suited to a given flooding problem. To reach an accord on the best methodological practice, it will be necessary to understand exactly what the flood models are doing. This could be facilitated through the use of the "virtual laboratories" approach proposed by Ceola et al., 2015.

2.2. Flood modelling: theory and application

The modelling of floods requires the derivation of methods which can simulate the flow dynamics of water. Most often this involves a 1D routing scheme, or the use of a 2D or 3D distributed hydrodynamic models (Teng et al., 2017). These schemes are derived from fluid mechanics, which is based around the application of Newtonian mechanics and calculus to describe a fluid. In this section, I will discuss the salient aspects of the theory of fluid mechanics, which lead to the possibility of inundation modelling, and therefore the growing field of global flood modelling.

2.2.1. Mechanics of water – progress to Navier-Stokes equations

The history of fluid mechanics can be seen as a history of describing the phenomena of fluids and converging on principles which can describe the full range of its behaviour. Scholarship into the behaviours of fluids, specifically water, has a long history. The principle of buoyancy, as described by Archimedes, is such an example, where the property of fluid density as a central component was introduced (Encyclopedia Britannica, 2014). In 1663, Pascal formalised these principles, which had been developing over time. With his thesis "On the equilibrium of fluids", Archimedes' principle was incorporated into the science of hydrostatics (Pascal, 1663).

A landmark moment for the development of fluid mechanics came with the work of Isaac Newton. Indeed, the conception of fluid mechanics is predicated upon the use of Newtonian mechanics, and the application of his famous laws to fluid masses (Newton, 1687). From these notions of water's behaviour as fundamentally mechanical, d'Alembert described the motions of fluids as wholly relatable to a dynamic equilibrium which they possess, and furthermore, individual strata of water within the body also share this equilibrium (Encyclopedia Britannica, 2017). These equilibria were formally resolved through the mathematics of Leonhard Euler,

whose work on partial differential equations allowed the derivation of clear equations to represent and unify these and other theories (Euler, 1768).

The culmination of fluid dynamics and its descriptive function reaches a peak with the Navier-Stokes equations, of which there are many other considerations besides what was briefly described above. The Navier-Stokes equations are equations which fully describe the motion of viscous fluids (Fefferman, 2006). They are also exceedingly complicated, as they have fully generalised the behaviour of fluids across all scales. However, for many contemporary purposes, which includes flood inundation modelling, many of the details are unnecessary, meaning several assumptions can be made to reduce their complexity.

2.2.2. Decomposed to Saint-Venant's equations (mass and momentum)

Specific engineering applications of the Navier-Stokes equations to water bodies like reservoirs, rivers and coasts means that decomposed forms of Navier-Stokes are used. This is because in many circumstances all the terms of Navier-Stokes are not necessary to obtain a sufficient representation of the fluid for the intended application. As a result, it is much more common to work with simpler mass and momentum equations, which can be solved in 1, 2 or 3 dimensions. In many cases, these simplifications involve reducing the various terms of the Navier-Stokes, such as force vectors, to the value of zero, as it is assumed that they have a negligible effect for the chosen application. This means we are left with equations which only account for gravitational and inertial forces, for example. For simplicity, the 1D equations will be shown here, themselves being used regularly in many modern applications, such as in the modelling tool HEC-RAS (USACE, 2010). The mass equation, when applied to open channel flow, can be expressed as:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0$$

Where A is the cross-sectional area at location x and time t, and Q is the discharge of water at location x and time t. Note the use of partial differential notation, as well as the zero term on the right hand side, describing what d'Alembert discussed as the dynamic equilibrium of fluids (Encyclopedia Britannica, 2017). Corresponding to this, the momentum equation for 1D open channel unsteady, non-uniform flow can be expressed as:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + g\frac{\partial h}{\partial x} + g\left(S_f - S\right) = 0$$

Where *u* is the dimensional velocity of flow, *g* is the acceleration due to gravity, *h* is the depth of the water at location *x* and time *t*, S_f is the friction slope, characterised by shearing effects of the channel perimeter, while *S* is the bed slope. Each term represents an external forcing acting on the motion of the water body. $\frac{\partial u}{\partial t}$ represents the local acceleration term, $u\frac{\partial u}{\partial x}$ represents the advective acceleration, a results of changing river geometries and $g\frac{\partial h}{\partial x}$ is the external force due to the effects of pressure. (Saint-Venant, 1871).

These equations have a high degree of flexibility and form the basis for almost all of the contemporary flood modelling software in use (Teng et al., 2017). Inclusion of the various terms described in the equation above leads to a corresponding difference in whether the effects of a dynamic, diffusive or kinematic wave is being modelled.

2.2.3. Development of numerical modelling schemes

While the principles of flood mechanics are based on the continuous nature of flood flow, actual computation of flows requires a process of discretisation of a modelling domain, which smaller portions having some relationship in space and time. Hence the differential term of ∂x as a continuous term becomes Δx, as the incremental increase must be defined for computation, rather than being infinitesimal in theory. This move to incremental increases and decreases is the process of discretisation.

Discretisation of a domain in flood modelling generally takes three approaches: finite difference, finite element and finite volume. The simplest format for inundation models is finite difference, which simply calculates the change in discharge *Q* and depth *h* across each cell in space and time. Attention must be given in that certain applications of the finite difference method can produce instability, and the Courant number, a dimensionless number which is a measure of instability, must be adhered to (Bates et al., 2010). Its definition is:

$$C = u\frac{\Delta x}{\Delta t}$$

Where "u" is the characteristic speed of the process being modelled, Δx is the spatial step in the model and Δt is the timestep. The further methods of finite element and finite volume are generally more computationally intense, and require definition of the water body into specific elements, the interaction of which is calculated in this numerical scheme. In particular, the use of finite volume method is more appropriate to specialised problems which

cause traditional methods to become unstable. Such an example is the dam-break problem (Valiani et al., 2002), which due to the discontinuity of flow is highly difficult to model.

2.2.4. Modern practices and applications

Just as the Saint-Venant equations can be described in up to three dimensions, so there are models which run numerical simulations of each these dimensions. Within each dimension, there are terms which can be excluded, therefore reducing the complexity of the model and its ability to fully describe the behaviour of the water, but which means the computations can occur much more quickly. For example, the model HEC-RAS is defined and runs 1D simulations of flows, and the number of terms included can be selected by the user. At the other end of the spectrum, full 3D hydrodynamic modelling is done with software such as TELEMAC 3D, allowing for the modelling of complex coastal and estuarine environments, and the mixture and dispersion of various pollutants within (Bedri et al., 2011).

For the applications of flooding, a very common setup of the model is to use 2D formulations of the Saint-Venant equations, given that overland flow is normally satisfied by the expressions of the 2D shallow water equations. There are many incumbent models which work in this arena, such as DIVAST (Falconer, 1986), MIKE21 (Warren & Bach, 1992), TUFLOW (WBM, 2001) and LISFLOOD-FP (Bates & De Roo, 2000). Given that the Bristol GFM referenced throughout this study uses LISFLOOD-FP as its hydrodynamic component, this will be described in brief and used as the example for the following section on modelling requirements.

As previously mentioned, the number of terms used from the Saint-Venant equation constitutes a modelling choice. In the case of LISFLOOD-FP, that modelling choice is to use all the components of the momentum equation besides the convective acceleration term, this term being omitted as it is related to high Froude conditions and therefore not important on many relatively flat floodplains with gradually varying flows (Bates et al., 2010):

$$\frac{\partial u}{\partial t} + g\frac{\partial h}{\partial x} + g(S + S_f) = 0$$

Which is discretised over a raster DEM. The benefit of such an approach means that the actual computation of an entire event of a domain of a reasonable size can occur quicker than a scheme that solves the full shallow water equations.

2.2.5. Data requirements

This section will cover the data requirements for flood inundation modelling. The application of flood models is computationally demanding and requires ancillary data to correctly model the extent of flood flows. Without accurate delineation of the environment which is to be modelled, it will not be the case that the mathematical descriptions of water will be enough to make accurate predictions (Beven, 2007; Saltelli et al., 2007).

2.2.5.1. DEMs

The use of digital elevation models (DEMs) is an integral component of inundation modelling (Bates & De Roo, 2000). Delineation of the topography in a digital domain is necessary for accurate prediction of flooded extent and simulating floodplain dynamics. The resolution of DEMs has been shown to be very important, as in the case of much coarser grids, important hydraulic features of a watershed can be missed. For example, in the 90m Shuttle Radar Topography Mission DEM, it was found that 96% of the channels assessed in the Amazon basin are not represented (Trigg et al., 2012). The impact of not representing these smaller channels is that stream connectivity and backwater effects onto the floodplain are not appropriately accounted for, having an adverse effect on simulation performance (Mateo et al., 2017). While it is broadly supported that finer resolution DEMs are better able to represent flood response, it is also claimed there is often a lower limit beyond which the extra computational cost has marginal additional benefits below a range of 50-120m (Jarihani et al., 2015; Savage et al., 2016a).

Most DEM models are derived using remote sensing and vary greatly in cost, resolution and performance. The highest resolution DEMs are those acquired with airborne LiDAR, which is able to provide 0.25-2m resolution and can be used for very localised detailed studies (Savage et al., 2016b; Schumann et al., 2014; Teng et al., 2015). Alternatively, there are much coarser scale DEM products, such as SRTM, ASTER and MERIT with global coverage, which are utilised for the modelling of data-scarce basins (Yan et al., 2015, Yamazaki et al., 2017). These DEM products are available at much coarser resolutions, between 100m to 1km, with errors in the SRTM being reported as approximately 8m (Rodríguez et al., 2006), while MERIT improved this coverage with up to ±2m accuracy from 39% to 58% (Yamazaki et al., 2017). There is an area of debate around their effectiveness given the typical uncertainties in the data, such as the imprecisions of the SRTM (Rodríguez et al., 2006), as well as the concern that with the

aging of the datasets, they are beginning to represent features which are no longer present in a rapidly developing world (Sampson et al., 2016). Indeed, over the past 30 years, 90,000km^2 of surface water has been displaced, while more surprisingly, approximately 184,000km^2 of extra surface water has been added to the water surface – mostly due to reservoirs (Pekel et al., 2016). The implications that this has at the regional level for inundation modelling are that there can be cases where the information used for represent surface water attenuation is now incorrect, presenting a further challenge to the modeller.

2.2.5.2. Bathymetry

Bathymetry is measure of the depth and shape of the channel bed. It determines local routing dynamics, as well as augmenting the effects of friction (Bates et al., 2010; Pappenberger et al., 2008). While it is arguably a sub-set of the DEM data, it is often not actively derived from the data, although in specialist studies remote sensing has aided in the creation of the bathymetry with more standard bed-profiling techniques, such as sonar and cross-section survey (Domeneghetti, 2016). This however is rarely accessible, meaning that it is often treated as separate from the DEM. As a result, it is often the case that assumptions have to be made about the shape of the channel, such as it is rectangular in shape (Bates & De Roo, 2000). By contrast, another strategy has been implemented to utilise alternative information within the data, such as discharge, to then calibrate the bathymetry *a posteriori*, exploiting established relations of hydraulic geometry (Leopold & Maddock, 1953; Neal et al., 2015; Wood et al., 2016). This is explored in more detail in chapter 3, as a variant of this approach is applied to the Bristol GFM.

There are also many positive side effects of correct estimation of channel geometry. In the estimation of flood frequency analysis, when there is little information available, it has been demonstrated that channel-geometry relationships help inform the magnitude of flood discharges in ungauged basins (Wharton & Tomlinson, 1999). In the case of 1D flow routing equations it is also highly important as the computation of flow propagation is dependent on river cross-sections, which themselves are derivatives of bathymetry (Yamazaki et al., 2011). Furthermore, accuracy in bathymetry has an important knock-on effect for river ratings (Coxon et al., 2015; Di Baldassarre & Montanari, 2009), often a valuable resource when concerning the estimation of boundary conditions within the model.

2.2.5.3. Boundary conditions

Boundary conditions are an essential part of the modelling scheme, and dictate the amount of water coming in and out of the modelled domain. Upstream boundary conditions are often the source of flow in fluvial models. Here the boundary condition would be the input hydrograph, dictating the peak flow, as well as the total water volume entering the domain over time. It has been found that accurate estimation and timing of the peak is critical when simulating past events (Grimaldi et al., 2019; Zhao et al., 2017). Downstream boundary conditions can be the source of flooding, particularly when considering a coastal flood model. In such circumstances the effects of storm surges or tidal mechanics can be represented as a downstream input discharge or more often a water level (Nguyen & de Kok, 2007).

2.2.5.4. State variables and parameters

State variables and parameters can represent aspects of the system which require estimation or modelling, but are often not known precisely, and in the case of state variables can change over the course of the model simulation. Such a state variable can normally be soil-moisture. While this has been shown as a dominant mechanism control in the case of flooding (Berghuijs et al., 2016), it is often assumed that the soil-moisture of the domain is already at peak capacity, hence allowing the modeller to proceed under the assumption that the land-surface is an impermeable membrane (Bates & De Roo, 2000). Indeed, it has been highlighted that this assumption, while convenient, is reducing the potential of global modelling (Grimaldi et al., 2019).

Parameters are components of the model used to represent processes which can't be represented in a more complex manner, for reasons which can either relate to computational cost or a conceptual gap in knowledge (Beven, 2002). A central parameter across all inundation modelling studies is the Manning's roughness coefficient. This is widespread across flood studies and in scalar form represents the energy loss due to roughness of channels and floodplains in the environment. As such, the Manning's *n* is consistently found to be an important control on inundation, and its estimation has to be carefully considered (Dimitriadis et al., 2016; Hall et al., 2005; Koks et al., 2015; Pappenberger et al., 2008).

2.2.6. Data-driven flood hazard modelling methods

Despite the prevalence of physically based approaches to flood modelling and the proliferation of datasets (Musa et al., 2015) with the potential for global-scale modelling (Lin et al., 2016; Sanyal & Lu, 2004; Yan et al., 2015), there are also a variety of competing methodologies to provide approximations of flood inundation, based on simple analyses of morphology.

2.2.6.1. Geomorphic estimations from DEM's

The use of topography to identify flood prone areas is not a recent development, as the use of the topographic wetness index has been shown to be a useful proxy of areas at risk of inundation, against a suite of other indicators (Franks et al., 1998; Samela et al., 2017). The topographic wetness index is an indicator which determines the susceptibility for flooding at a location given the surrounding terrain. This is derived from upstream area and the local slope. The original topographic wetness index is defined as:

$$TWI = \ln \frac{a}{\tan b}$$

Where 'a' is the upstream area travelling through a point per unit contour length, and 'b' is the local slope, in radians. The resulting value of TWI is a dimensionless number, which gives an indication of the susceptibility of flooding, at the given location. As it incorporates upstream area and local slope, it is able to crudely represent the changes of water pooling and "inundating".

Further studies have found that use of remotely sensed DEMs is viable for the calculation of flood prone areas and that SRTM has a consistent reliability for the task (Manfreda et al., 2011), underlining the fact that such methods can be quickly employed for rough delineation of flood risk, and could be used to cede prior estimations of parameter values in ungauged basins.

2.3. The development of global flood modelling as a practice

Along with the developments outlined above, there are several other recent developments which have allowed for the transition from local-scale flood studies to the global scale. The emergence of global flood modelling is now established (Schumann et al., 2018), which has grown out of the preceding ideas of considering the hydrological sciences in a global perspective (Eagleson, 1986). It is worth noting these developments, as they are to inform

the necessary research aims, and align with the validation frameworks which have already been recommended.

2.3.1. Proliferation of data sources with global coverage

Recent advances in remote sensing now allow for the modelling of much larger continental-scale domains. Global DEM's such as SRTM are behind this development, and have been noted as being valuable for continental scale inundation modelling (Lin et al., 2016a; Musa et al., 2015; Sanyal & Lu, 2004). This is particularly relevant to the question of modelling in a global context, as so often the great obstacle to such endeavours is the scarcity or complete lack of data (Salinas et al., 2013). Other global datasets of great value to hydraulic modelling are global river width databases, with a corresponding analysis which assesses the global relations of river width to a number of other integral hydrological variables (Frasson et al., 2019; Yamazaki et al., 2014a). Also noted as being very important for the continued development of global flood modelling is addressing the sensitivity of global models to assumptions about local flood defences (de Moel et al., 2015; Sampson et al., 2015; Ward et al., 2013). However, with the release of FLOPROS, an open source and growing database of flood protection standards, it becomes possible to present some level of defence representation in a consistent manner, which will continue to improve (Scussolini et al., 2016).

Global hydrology is also beginning to be better understood, which presents the opportunity for us to accurately condition the boundary conditions for flood models. Particularly in the case of Europe, there are well calibrated hydrological models which are able to represent the water cycle and highlight trends (Hundecha et al., 2016). However, there is still some way to go before GHMs can produce high resolution hydrographs for peak discharges (Sood & Smakhtin, 2015). However, given the importance of river routing for the derivation of peak discharges (Zhao et al., 2017), the value of global hydrological models for extreme events might be limited. The effectiveness of GHMs has been seen more in the field of accurate water balance modelling and more general hydrologic signatures, where it is typical, repeatable hydrological behaviour. In contrast, extreme events are generally atypical, therefore meaning that models which are focussed on the general water balance at the global scale may offer little to the task of peak flow estimation.

2.3.2. Greater computational resource

As models increase in size, the computational demand also increases. Hence for global models, and the 2D solvers, there are upper limits to the size of domains which can be modelled within a practical time frame with current computing hardware (Toombes & Chanson, 2011). On a single core, the amount of RAM available will determine how many cells can be computed in unison, while the CPU determines the speed. The CPU provides the computational power available to the modeller for the simulation of hydrodynamics and while the RAM does slowly increase, an area of exponentially greater gains is using parallelisation and high-performance computing. These innovations allow for model domains of a larger size to be modelled. However, this does create a new question of what should be modelled on which core.

2.3.3. Parameterisations which incorporate remotely sensed data

The proliferation of remote sensing products can genuinely be thought of as a "flood of data". However, the correct way to utilise this data is still a matter for debate and it has been openly identified that continuous spatiotemporal datasets are not suitable replacements for genuine understanding of continuous spatiotemporal processes (Di Baldassarre & Uhlenbrook, 2012). As a lot of these new datasets, such as the DEMs, are at a coarser scale of resolution, it means there is a representational gap between the data and the geographical features which wasn't encountered at the local level. This constitutes a knowledge gap about the natural system, which can be inferred through the coarse data, but essentially requires an assumption on the part of the modeller (Beven, 2002). This can be as fundamental as the assumptions about the channel network being modelled. Many of the GFMs are reliant on a topographically-derived channel network, which is distinct from knowing the exact network (Lehner et al., 2008). In ungauged basins, the accuracy of these networks is extremely difficult to validate and not possible in practical terms, meaning it could be possible to over- or underestimate the extent of the river network. Therefore, if considerations are not made, then it can be easy to derive false confidence about the value of the modelling information derived as in the case that floodplain connectivity is wrongly estimated, hence the need for novel approaches and parameterisations within the GFM community which accommodate these situations (Dottori et al., 2013).

Undoubtedly one of the principle parameterisations which facilitates global flood modelling is that of channel subgrid parameterisation (Grimaldi et al., 2019; Schumann et al., 2016). Subgrid parameterisation ensures rivers narrower than the cell size aren't removed from the model. Due to the limitations of SRTM, with its finest resolution being ~90m, it often misses a large number of smaller tributaries, as in the case of the Amazon (Trigg et al., 2012). In many of the current suite of GFMs, subgrid parametrisation is a necessary component (Sampson et al., 2015; Siqueira et al., 2018; Yamazaki et al., 2011). The Bristol GFM uses the subgrid parameterisation algorithm present in LISFLOOD-FP (Neal et al., 2012). When LISFLOOD-FP was used to model an "ungauged basin" at a domain resolution of 5km, it was still possible to achieve a fit of 64% using the critical success index metric, covered in more detail in chapter 4 (Komi et al., 2017), indicating that the concept behind the parameterisation scheme is able to capture fundamental channel-floodplain processes at large scale (Clark et al., 2017). Channel networks and connectivity can often relate to the process of channel bifurcation. This relates directly to the capacity of channel networks during flood events and it has been demonstrated that when incorporated an improvement to performance is seen (Eilander et al., 2018; Yamazaki et al., 2014b).

The global perspective of inundation modelling will struggle to be effective if regional and local components are overlooked (de Moel et al., 2015). As subgrid parameterisation is so prominent, it will be important not to overlook the morphological components of the river systems (Neal et al., 2015). A key morphological indicator of the river system is the bankfull discharge (Arbeláez & Posada, 2007), which itself can be used to help define the relations of hydraulic geometry (Castro & Jackson, 2001). The bankfull discharge is the return interval of how often the river flow will break the banks of the channel (Williams, 1978). Generally speaking, the bankfull discharge interval resides somewhere between a value of 1.5 to 2 years (Ahilan et al., 2013; Castro & Jackson, 2001), while rivers in more porous environments can have an interval of 2 to 3 years (Petit & Pauquet, 1997), and certain arid environments have an interval measured in decades (Ahilan et al., 2013; Williams, 1978). The feedbacks of this variable to different aspects of the river system highlight the necessity of the careful estimation of these parameters in the GFM methodologies.

2.3.4. The contemporary global flood modelling community

Of the prominent GFM studies conducted so far (Bernhofen et al., 2018; Trigg et al., 2016), six models and their respective outputs have been assessed, as shown in Table 1. These models were all published within five years of each other, testifying to the remarkable speed at which the subfield of global flood modelling has developed, and highlighting the antecedents needed to conduct such research. The earliest of these is the 1D diffusive wave routing methodology with floodplain storage cells of CaMa-Flood (Yamazaki et al., 2011), generated river flooding through the global simulation of river flows which were conditioned on Global Runoff Data Centre (GRDC). A year later, a methodology for deriving hazard maps of fluvial flooding was published from the ECMWF (Pappenberger et al., 2012), which again makes use of a river routing scheme, but derives peak forcing's from climate data sets and an integrated land surface model. In 2013, the framework for GLOFRIS was published which coupled a hydrological model PCR-GLOBWB to a dynamic routing scheme. This model focusses more specifically on the risk aspect, explicitly incorporating losses into the framework, and drives the hazard component of the Aqueduct Flood Risk Analyzer (Winsemius et al., 2013). In 2015, two further models were published, the CIMA-UNEP model, and the model of this study: The Fathom/ Bristol GFM. The first model derives peak flows from a regionalisation scheme based off 8000 gauges from the GRDC. This scheme is further conditioned with the use of a global hydrological model to improve its suitability for ungauged basins. Again, the hydraulics are derived from a 1D routing scheme (Rudari et al., 2015). The Bristol GFM meanwhile generates peak flows using a regionalised flood frequency analysis, which produces peak flows for a given return period. These flows are used as forcing's for the underlying 1D-2D hydrodynamic model of LISFLOOD-FP with a subgrid scheme (Neal et al., 2012), which can then produce hazard maps (Sampson et al., 2015). The final GFM published in 2016 is from the Joint Research Centre. Making use of the hydrological modelling outputs of the Global Flood Awareness System (GloFAS) (Alfieri et al., 2013), these outputs are then used as forcing's into a hydrodynamic model, which is again LISFLOOD-FP. The outputs of these simulations are then merged to create flood hazard maps of a given return period.

GFM	Peak flow estimation	Inundation approach	Resolution	Reference
Bristol GFM	Regional flood frequency analysis to drive flows for hydrodynamic model	1D-2D hydrodynamic model	90m	(Sampson et al., 2015)
CaMa-Flood	Climate and water balance modelling to drive global river routing	1D routing hydraulic model	540m	(Yamazaki et al., 2011)
GLOFRIS	Climate and hydrological forcing for floodplain storage modelling	Global 1D routing	900m	(Winsemius et al., 2013)
Cima-Unep	Regional flood frequency analysis with hydrological modelling and derived flow-elevation relationships	1D hydraulic model	90m	(Rudari et al., 2015)
JRC	Climate analysis and land-surface modelling used to force hydrodynamic model	2D hydrodynamic model	900m	(Alfieri et al., 2013)
ECMWF	Climate analysis and land-surface modelling used as a driver for storage modelling	1D river routing	540m	(Pappenberger et al., 2012)

Table 3. 1 - GFMs currently used for cross-model study.

These developments represent a large variety of different assumptions and processes by which flood hazard is simulated on the continental scale. Indeed, the only real commonality between all of them is this commitment to a "from data to hazard" methodology. By this I mean the entirety of the flooding process is derived from data. There is a 3-step process that each of the models perform. The first step is the generation of peak flows – either using regionalisation of gauge data, hydrological models, or a combination of both. The second step

is the simulation of inundation from these predicted flows in every reach or basin, which in the final step are combined as a hazard map with complete coverage. However, these are not the only models to follow this route. Although the publications on each model are ostensibly *global* flood model focused, there are similar studies which model at the reach to continental scale. For example, the coupling of hydrological and hydraulic modelling to simulate the Ob river in Siberia (Biancamaria et al., 2009). Similarly, there are efforts to simulate the entirety of South America through a coupled hydrologic-hydrodynamic model, which can also simulate flooding (Siqueira et al., 2018), although both of these studies fall short of quantifying the hazard with the coverage of the global models. While encouraging, it has been identified that there are a number of improvements which such GFMs need to make (Grimaldi et al., 2019; Vorogushyn et al., 2018). Some of the suggestions made are finding alternative streams of information to allow better decisions to be made, leading to call for denser monitoring networks in data-sparse areas. Another clear area for improvements in GFMs is one which has been identified as important in flooding and hydrological sciences for decades, that of uncertainty (Ward et al., 2015).

2.4. Uncertainty in flood predictions

It is clear from the preceding discussions that the continued development of GFMs requires consideration of core data and model uncertainties. It has already been briefly touched upon how the transition from continuous phenomena to discrete descriptions thereof can be a source of uncertainty (Beven, 2002). Not only in global flood modelling, but in the broader fields of flood risk (Levy & Hall, 2005) and hydrology (Montanari, 2011), there is a wealth of literature on the importance of considering uncertainty. Alongside this is the development of the methodology in this thesis, which is attempting to understand the behaviour of GFMs under uncertainty using primarily sensitivity analysis techniques. While current calls to implement a more rigorous framework for GFM model comparisons are definitely needed, it remains the case that the discussions about the role of uncertainty are under-represented (Hoch & Trigg, 2019).

2.4.1. What do we mean by uncertainty?

The field of model uncertainty assessment in the hydrological sciences is vast. As far back as 1969, hydrological research was preoccupied with the problems of parameter identification, sensitivity analysis and the relations of model inputs to model outputs (Vemuri et al., 1969).

Yet despite this, there have still been discussions and debates about the meaning of uncertainty 40 years after the issue first came to prominence (Montanari, 2007). One reason for this confusion among the community has been a philosophical disagreement (in a statistical sense) about how uncertainty can be parsed mathematically – a debate of frequentist against Bayesian thinking (Montanari et al., 2009). According to Nearing et al., 2016, this dichotomy of Bayesian against frequentist thinking is actually an artefact of incorrectly defined philosophical terminology, which creates the illusion of a disagreement that doesn't exist. Instead, they convincingly argue that "*apparent limitations of probability theory [as applied in uncertainty analysis] are not actually consequences of that theory, but rather of deeper underlying epistemological (and ontological) issues.*". Despite this, there are interesting perspectives to be had from consideration of axiomatic difference, so a suite of uncertainty-based definitions will be considered (Nearing et al., 2016).

2.4.1.1. Definitions

"*What makes modelling and scientific inquiry in general so painful is uncertainty. Uncertainty is not an accident of the scientific method, but its substance."* (Saltelli et al., 2007)

When defining the complementary analyses of sensitivity and uncertainty analysis, Saltelli et al., 2007 state that sensitivity analysis is concerned with how "uncertainty in the output of a model can be apportioned to different sources of uncertainty in the input". Consequently, the purpose of the uncertainty analysis is to quantify the uncertainty of the output, which by extension means quantifying the uncertainty of the input. In this distinction we can see that the purpose of sensitivity analysis is entirely preoccupied with the internal behaviour of the model, but it is in uncertainty analysis that the interaction of the model with information about natural systems is recognised.

2.4.1.2. Sources of uncertainty

According to Götzinger & Bárdossy (2008), sources of uncertainty are usually separated into four categories to aid their definition and estimation, and are widely cited in the literature. The first of these is input uncertainty. This is related to the imperfect knowledge and measurement of the input values, the likely not ideal location of the data, etc. In the case of DEM measurements for example, this uncertainty is related to the expected deviance from the actual elevation, which in the case of SRTM averages 8m (Rodríguez et al., 2006). Another uncertainty source comes in the form of state variables, which are variables in the model

which represents a part of the natural system that can change over time, but are neither a forcing or an output. A possible example of a state variable is something like filled reservoir volume, which necessitates the initialising or "warming up" of a model (Kim et al., 2018). Such behaviour is normally needed as state variables are generally autocorrelated and possess "memory" in simulations (Ammann et al., 2018). A further source of uncertainty comes about from the encoding of natural processes using equations. This is defined as "process abstraction-related uncertainty". A very common equation used in open channel modelling is the Manning's equation, which uses a few key parameters as a proxy for the flow. The corresponding parameters of these equations also fall within this uncertainty type, more generally known as parameter uncertainty. Hence in the case of the Manning's equation, the parameter Manning's n is most often the source of uncertainty, as all other parameters in the equation are measurable. The final source of input uncertainty is that of model structural uncertainty.

The other usual distinction which is made about uncertainty is that comes as either epistemic or aleatory uncertainty. Clear definitions are given by Beven et al., 2011, who states that epistemic uncertainty is the variance of a system which can't be described by a stochastic variation. This means an uncertainty which is based simply on a lack of clear knowledge, as described (although not exclusively) by model structural uncertainty and the encoding of natural processes in mathematical form. By contrast, aleatory uncertainty is understood as the apparent random variation of a variable or input (Savage et al., 2016a), hence its easier description by probability distributions, as in the case of sensitivity analyses (Weichel et al., 2007). This ease comes about by it being considered as a variation which is simply natural, meaning that it is unable to be reduced with more information or computation.

2.4.1.3. Redefining uncertainty as information

Contrasting with the traditional forms of uncertainty given as primary sources, such as parameter uncertainty, model structure etc. (Götzinger & Bárdossy, 2008; Liu & Gupta, 2007); Nearing et al., 2016 instead presents these forms as "proximal sources" of uncertainty. This definition is derived from their wider discussion about how the assumptions underlying current modelling philosophies are lacking some of the necessary logic. Instead, there are underlying "fundamental sources" of uncertainty to address first, from which it logically follows that the "traditional" sources of quantified uncertainty are secondary. The first

fundamental source is a problem of finitude, of which there are two cases. In the first case, there is the finite experiments problem; this being that only a finite number of experiments are possible. As a consequence, it is only possible to falsify hypotheses made, rather than it being possible to make a truth statement (Popper, 1963). The second case is that of finite hypotheses, meaning that there is only the option of testing a subset of all possible models, i.e. descriptions of reality which interests us, or that are the only ones available to us. In effect, what this means is that "*the act of not testing some potentially correct model is indistinguishable from assigning to that model zero probability* in the inference prior" (Nearing et al., 2016).

The final source of fundamental uncertainty described is the Duhem-Quine thesis (Harding, 1975), which states that in any experiment, it is impossible to test a *single hypothesis*, because within that single hypothesis is a number of implicit assumptions about the natural system which mean the hypothesis can't be considered in isolation. In the case of hydrological and inundation modelling, that presents itself as a problem because the complex systems of catchments have interacting physical processes across multiple scales (Wagener et al., 2007). Therefore, through the applications of Monte Carlo simulation, we can say that it is possible to have thousands of working hypotheses about the model, if one of these should be true, we are left with the situation that this specific truth is only understood through the bundle of assumptions carried within the model setup, meaning that we can't ascribe any higher value to models than that of being a heuristic (Oreskes et al., 1994). Hence the common usage of the aphorism "all models are wrong, some are useful" (Box, 1976).

2.4.1.4 Present the viewpoint of space-mapping and pragmatic modelling

Despite the reductionist view of the value of models described above, it is still held that they can be useful. However, it is recommended that the idea of uncertainty is instead conceptualised as information, as in "*how much information do we have, and how do we use it?"* (page 1, Nearing et al., 2016). In this perspective, the phenomenological limitations of modelling are acknowledged. From this, it can be pragmatically realised that the representation of hydrological systems in models serves a specific purpose, rather than being able to validate any particular theory (Giere, 2004). At this point, it could be asked whether this high-level discussion adds any value to the task at hand. My response to that is given that global modelling is a new field, which as mentioned operates on the basis of "from data to

hazard", it is appropriate to incorporate advanced thinking on uncertainty into the discipline from its inception, as there are surely a lot of implicit assumptions therein, meaning the community should be careful to not derive 'broad-stroke' trends about the state of flooding worldwide given the results of one of the models.

One of the pragmatic concepts that could realise these aims is that of *space-mapping* (Beven, 2002). By this, it is meant the modelling of the natural systems is understood as a mapping of the landscape space *out there*, and the model space internally defined by us. In the concept of *landscape space* there is already the implicit assumption that we are unable to perfectly describe reality, meaning by extension that we are unable to perfectly inscribe reality into our modelling setup. Hence the mapping of landscape space to model space as an inherently fuzzy component, which as a modeller, we can accept as a reality of the model, its "substance" (Saltelli et al., 2007). However, in such a scheme of mapping, the effects of uniqueness of place are still given their due in landscape space (Beven, 2000), but in mapping to model space, this fuzziness allows for the ideas of similarity of catchment processes to be realised without apparent contradiction between similarity and uniqueness (Gupta et al., 2014).

2.4.2. Review of uncertainty within hydrological sciences and flood inundation

Although newly equipped with a fresh perspective on uncertainty, it remains the case that in application, the common methods remain the same – that is, sampling and Monte Carlo simulations. Therefore, the ideas discussed above will be used to help interpret the results, rather be used to formulate an alternative Monte Carlo method.

There have been studies which have addressed the problem of correctly representing the important features of a landscape. In the case of larger scale flood models, the morphological features of the model have been identified as highly important (Grimaldi et al., 2019), while the effects of DEM uncertainties due to the elevation (Hawker et al., 2018; Wilson & Atkinson, 2003) and cell size (Savage et al., 2016b) have been well established. Another area which has received attention has been in the estimation of peak flows or the analysis of flood frequencies (Parkes & Demeritt, 2016). One such area is on the derivation of peak flows by assuming different distributions for the flow data one has available, where it is found generalised extreme value (GEV) characterises peak flows adequately. However, the discussion by Hall & Anderson, 2002 notes that the paradigm for estimating peak flows can't be based purely on the relation of observations to model, given the conditions of data

scarcity. Instead, they suggest other paradigms are used. An example of a possible alternative comes from the estimation of flood frequencies using continuous simulation at a catchments, allowing for the synthesis of more extreme values through which to drive hydrological extremes (Blazkova & Beven, 2009; Cameron et al., 2000a; Cameron et al., 2000b; Cameron et al., 1999).

The effect of channel bathymetry on peak flow has also been studied, given that there can be significant errors in flow estimation at even a well gauged river, having a subsequent effect for extreme value estimation (Di Baldassarre & Montanari, 2009). This also has an effect on the rating curves, of which the uncertainty has been quantified in many ways (Ocio et al., 2017). Domeneghetti et al., 2012 found that the uncertainty of rating curves can cause the miscalibration of parameter values if not correctly accounted for. This can be done, by assessing errors as possibly "epistemic", such as unanticipated artefacts in the channel (McMillan & Westerberg, 2015). This analysis was generalised further to a broader framework of quantifying rating curve uncertainty and was found to effectively parse the broad range of issues which can contribute to rating curve uncertainty (Coxon et al., 2015).

2.4.2.1. Implications for policy and risk

It is also worth remembering the reason for performing uncertainty analyses, that there are stakeholders who receive and make decisions based on the information inherent to flood risk models. There are big implications of uncertainty analysis for risk management because it has to be effectively communicated that predictions are not deterministic, and it is crucial for disaster risk reduction that the modelling community is able to do so (Grimaldi et al., 2019), alongside the calculating the appropriate investments to make in disaster risk reduction. Visualisations of risk as probabilistic, through the use of confidence intervals, can better explain the extent of uncertainties in outputs and inform stakeholders how to use resources to better refine uncertain predictions (Beven et al., 2015). Not only at the stakeholder level, but also at the level of the non-academic practitioner, it is important to ensure that the concepts of uncertainty are suitably ingrained into the methodologies of analysing risk (Beven et al., 2011).

2.4.3. Background to uncertainty: Bayes theorem

Techniques for how uncertainty is quantified come under the disciplines of sensitivity/ uncertainty analysis (Saltelli et al., 2007). From a high-level perspective, uncertainty can be

defined as Bayesian (Romanowicz et al., 1996). Bayesian probability theory is based on the supposition that claims about the probability of the world are inherently subjective, as they require an observer. As a theorem, it states that the conditional probability of an event which is determined from multiple variables, is itself a function of the probability of those events and the relation of probability between them. Described mathematically, the equation takes the following form:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

Where P(B|A) is the probability of B given A. P(A|B) is the probability of A given B. In modelling terms, this is what is known as the likelihood function. P(B) is the prior probability of event B occurring, while P(A) is the marginal probability of the event A occurring.

This concept has been used in many uncertainty analyses through the application of Monte Carlo sampling, where each input distribution to a model has a prior probability of importance. A given value on a parameter range has a probability attached to it. This probability is computed with the probability of good performance, and the likelihood of one being linked to another (Apel et al., 2004; Aronica et al., 2002).

2.4.3.1. Sensitivity Analysis

Sensitivity analysis (SA) is the process of determining how much of the output variance from a model can be attributed to parts of the input variance. In this, it has a lot of crossover with systems analysis. In the hydrological sciences, there has been healthy discussion on the role and importance of sensitivity analysis for the last 20 years (Song et al., 2015), with guidance on how to apply and use methods becoming established (Pianosi et al., 2016). However, alongside such a proliferation of methodologies and modes of application, there are correspondingly as many potential interpretations of what the analyses mean (Razavi & Gupta, 2015), highlighting the need to take care when using sensitivity analysis techniques, through understanding what information it is one wants to extract with their chosen method, as many of the methods in principle will give the same approximate answers (Wainwright et al., 2014).

A standard reason for the application of SA is when there is some concern that a model might be overparameterized, or overcomplicated (Schoups et al., 2008), indicating that the model

is non-unique and exhibits equifinality – the phenomenon of multiple model inputs returning an identical or near identical model output (Beven, 2006). In such cases, the problem for the SA practitioner is to find signatures of the model's processes (Gupta et al., 2008), meaning that unique components of the model can be identified, and not become lost under a cloud of indeterminacy (Wagener & Gupta, 2005). This should mean that even the effects such as the input distribution of the parameters on the outputs can be observed (Benke et al., 2008). For such purposes the recently proposed SA method of dynamic identifiability analysis is suitable (Wagener et al., 2003), which assesses the overall performance of the model in tandem with the information which can be derived from the behaviour of individual components. It has been highlighted that another crucial consideration for sensitivity analysis practitioners is making sure that their analyses have converged and are suitably robust (Sarrazin et al., 2016). However, this is not a simple matter as every combination of model and technique has a different criterion for this, given that the structure of the models and the behaviour of the parameters themselves condition what constitutes expected model behaviour and allowable deviations (Gan et al., 2014).

Given the complexity of the GFM methodologies, it seems quite likely that such analysis will be necessary for each of the models. However, methods such as dynamic identifiability analysis (DYNIA) imply the need for a highly dense sampling strategy, and it is the case that many of the GFMs will have a lot of parameters and are computationally expensive over large areas. To simplify the model a bit more roughly, a screening of the parameters first can be appropriate, which can be done with the Morris method (Morris, 1991). Once the parameters are screened, another common approach is variance based methods, which are able to quantify precisely how much of the output variance is attributable to each input variance (Saltelli et al., 2010). While each of these analyses are grounded in supposedly opposing theory, the recent development of another SA method called variogram analysis of a response surface (VARS) claims that the differences in approach of each preceding method is one of scale, and it is able to reproduce the same results of each method (Razavi & Gupta, 2016). The Morris method is explained further in chapters 4 and 5, where its implementation is discussed in greater detail.

Sensitivity analysis of GFMs is currently a limited area. The effects of a small subset of structural changes have been analysed on the CaMa-Flood model, highlight the importance

of river bifurcations (Eilander et al., 2018b), while a quantitative analysis performance on the JRC model was a one-at-a-time analysis, so not able to definitively explain model behaviours or structure. However, it did indicate that channel roughness was found as insensitive (Dottori et al., 2016), contrary to other sensitivity analysis studies which usually determine channel roughness to be highly important (Hall et al., 2005; Pappenberger et al., 2008).

There have been many examples of the sensitivity analysis of reach scale inundation models, which also account for the spatial distribution of sensitivities (Hall et al., 2005; Savage et al., 2016a). An in-depth sensitivity analysis was conducted on the 1D routing model HEC-RAS, to assess the viability of each method in the context of flood inundation (Pappenberger et al., 2008); however in this study, each SA method supposedly gave different rankings. This is slightly unusual, given the general trend that SA methods if implemented correctly will broadly be in agreement (Gan et al., 2014; Wainwright et al., 2014); therefore, it reinforces the fact that sensitivity analysis needs careful design and implementation.

2.4.3.2. GLUE

One of the most cited methods for quantifying uncertainty in environmental models is the method of generalised likelihood of uncertainty estimation (GLUE) (Beven & Binley, 1992). In this methodology, Monte Carlo simulation is used to sample which combination of model parameters (and structures) return the best performance. Indeed, the confidence intervals derived from GLUE has been termed "sensitivity envelops" as a more appropriate description (Montanari, 2007). A key tenet of the method is that *"within the limitations of a given model structure and errors in boundary conditions and field observations, different sets of parameters/models may be equally likely as simulators of a catchment"*. Further to this, not only can multiple sets be simulators, but they can be acceptable, or good performances of the catchment (Beven & Freer, 2001). This concept is termed equifinality (Beven, 2006), that there can be parameter sets applied in the model which will return the same output. This means that there is a specific emphasis in the GLUE methodology setting priors loosely, without making very strong assumptions about the uncertainties, as these are defined by the model behaviour instead (Renard et al., 2010).

There are many examples which highlight the success of the GLUE method. Freer et al., 1996 applied this technique to a hydrological model, demonstrating a range of parameters as acceptable, and how the incorporation of new data could be incorporated to constrain the

prediction limits, via Bayesian updating, while it has also been demonstrated that the uncertainty of flood extent predictions using binary information can be derived also (Aronica et al., 2002). Another strong application of the GLUE method is in continuous simulation for flood frequency analysis, where a 1000 year flood peak could be estimated on a hydrological model, calibrated with limited data and then using a stochastic rainfall generator (Cameron et al., 1999). Within the GLUE framework, the flexibility of extending these estimates to account for the effects of climate change was also studied, with satisfactory results (Cameron et al., 2000).

Because GLUE emphasises less the importance of defining formal priors, it has received some criticism from various corners of the community, which developed into a fairly strong debate around the correct way to statistically parse hydrological systems and models. Montanari, 2005 suggested that there are some unspoken implicit assumptions of the method which are problematic, and that it underpredicts the total uncertainty of hydrological systems. Others have gone a step further and claimed that the subjective selection of acceptable model behaviour constitutes a level of arbitrariness which is simply unsuitable, due to the ad hoc likelihood functions employed (Stedinger et al., 2008), while the lack of formal statistical inference has been decried as simply incoherent (Mantovan & Todini, 2006). Beven et al., 2008 responded to this by pointing out that the strong statistical inferences of formal Bayesian approaches will present a well defined distribution of uncertainty, but having been derived from an overly simplified error model in their case study, the subsequent parameter distribution, while well-defined, is false. The debate was rebalanced by Vrugt et al., 2009, who demonstrated that there is a strong equivalence between formal Bayesian methods and GLUE, while more recently the discussion by Nearing et al., 2016 argued quite convincingly that the supposed differences of GLUE with the more formal methods are actually a case of philosophical, rather than statistical incoherence, suggesting that some of the stronger criticisms levelled against the methodology of GLUE were derived from an incorrect understanding of the first principles which drives the method. In a retrospective paper, Beven & Binley, 2014 noted that the main difference of the informal method of GLUE to formal methods is that in GLUE it is more explicitly obvious when the models fail.

2.4.4. Regionalisation of models for prediction in ungauged basins

One reason for such a rigorous treatment of uncertainty relevant to global scale modelling is that it helps address the problem of predictions in ungauged basins (Hrachowitz et al., 2013). For the GFM, this is an important consideration as the GFM methodologies are defined with the purpose of predicting in ungauged basins. After a decade of research as the main theme of the IAHS, there have been significant advances within the field of hydrological sciences. One of the main methodological advances is the regionalisation of models (Merz & Blöschl, 2004) based on catchment similarity (Wagener et al., 2007). Such methods can help to characterise the streamflow of similar regions, which has implications for water resource management (Abimbola et al., 2017). An effective way to determine commonality between gauged and ungauged basins is with the use of hydrological signatures, which describe catchment behaviours as derivatives of specific processes. A common signature to regionalise is the flow duration curve (FDC) (Booker & Snelder, 2012). While useful, there is still evident uncertainty in their application, as their interaction with the other proximal sources of uncertainties in models is difficult to quantify (Yadav et al., 2007), and should the regionalisation schemes employed fail to recognise these, it can often be the case that the subsequent regressions are over-conditioned (Westerberg et al., 2016). It is perhaps not coincidental that at the end of this research decade the emergence of our GFMs began, as both research drives are grounded in the attempt to understand the broader hydrology of the planet.

2.5. Flood Frequency Analysis

Flood frequency analysis (FFA) is an “old-school” technique in the hydrological sciences. It is a method for predicting extreme flow magnitudes in “hydrologically homogenous” regions (Dalrymple, 1960). The idea of hydrological homogeneity is pervasive in hydrology, with the concept of specific hydrologic landscape units representing this discrete homogeneity of regions (Winter, 2001). Despite the strong case for uniqueness of place, from a pragmatic standpoint it is still necessary to assume homogeneity, particularly when the FFA is preceded by a regionalisation procedure. Recent methods for FFA have centred on attempting express the underlying hydrological processes in the catchments, such as capturing the base flow index using the geology (Dawdy et al., 2012), but still relying on statistical methods to extrapolate from relatively small pools of data (Engeland et al., 2005) and using innovative methods of incorporating historical flood information as well (Parkes & Demeritt, 2016).

2.5.1. Process representation

Hydrologic similarity has strong implications for FFA, as there are multiple causes to which a flood can be attributed, with the exact causes varying across multiple scales (Tarasova et al., 2019). The use of homogeneity measures has been demonstrated to constrain appropriate estimates of flood peaks, with smaller clusters of gauges returning better estimates (Kumar & Chatterjee, 2005). Ouarda et al., 2006 has noted that better quality datasets are allowing the seasonality of hydrological behaviour to be used to improve prediction of flood frequency, as its ability to represent catchment processes has been noted across multiple studies (Ahn & Palmer, 2016; Hailegeorgis & Alfredsen, 2017a). In particular, Hailegeorgis & Alfredsen, 2017 attempted a regionalised flood frequency analysis of Norwegian catchments, based on the use of seasonality, high flows and runoff response, using l-moments (Hosking & Wallis, 1997). While effective, there was nonetheless significant uncertainty in the result. The Bristol GFM which is the subject of this thesis derives its peak flows from a global-scale regionalised flood frequency analysis (Smith et al., 2015) – a derivative of the standard flood frequency analysis approach, for use in ungauged basins. This is covered in detail in chapters 3 and 7.

It is the case that uncertainty will be present in the use of regionalised flood frequency analysis (rFFA). However, it is possible to quantify the uncertainty and compare the relative effectiveness of various FFA methods (Renard et al., 2013), while incorporating local information. from a variety of sources. This can help to better parse uncertainty in the flood frequency analysis, a process which has been termed “flood frequency hydrology” (Viglione et al., 2013). An effective rFFA method for dealing with uncertainty is the index flood method, which reduces prediction error with the inclusion of more data (Jingyi & Hall, 2004). The index flood method is a standard procedure, where if one has access to a series of annual maxima, the mean annual flood (MAF) can be calculated, as well as a growth curve which correlates the growth of the flood with the frequency of its occurrence. A full description of the index flood method is given in section 3.3. The index flood method is in fact recommended for rFFA methods, in its ability to discriminate catchment response to extreme events, due to the distinction of the growth curve and MAF (Meigh, 1995). A clear conclusion of its use has shown that representation of the ‘peakiness’ and intensity of flood behaviour in arid regions is possible and even effective (Farquharson et al., 1992), despite the problems of data sparsity associated with arid regions (Zaman et al., 2012). By contrast, the conclusions from rFFA

methods in the tropics and sub-tropics is that flood peaks are generally predicted as much less intense in growth, but the average magnitude is generally higher (Meigh et al., 1997).

2.5.2. Climate typologies and hydroclimatic indices

As a statistical method, the effectiveness of flood frequency analysis is wholly dependent on the type and quality of the data given. As will be discussed in chapter 3, the regionalised flood frequency analysis method used in this thesis is dependent on climate to distinguish relevant clusters of gauged catchments. While catchment size and rainfall are common predictor variables, at larger scales it is necessary to incorporate climate typology into the method, as this is a clear driver of hydrological regimes (van Werkhoven et al., 2008). Further to this, catchments at a more northern latitude have a greater sensitivity to changes in climate, with respect to their precipitation regimes, as the transition of rain to snowfall and the subsequent catchment response is governed by small temperature changes (Carey et al., 2010). As such, there is the pressing question of how climate should be best represented. Currently, the standard in hydrology is to classify based on the Koppen Geiger classification system (Peel et al., 2006), despite the primary indicators of its climate being due to vegetation. While this system has been effective, it does not reflect contemporary interest in the representation of hydrological processes. A viable alternative climatic classification system could be that described by Knoben et al., 2018, which describes the global hydroclimate in terms of three indices which reside on a continuous spectrum. These indices are aridity, aridity seasonality and percentage of precipitation as snowfall, found to be independent of each other. Within such a classification scheme, regionalisation effects have the potential to be more effective than in the case of Koppen-Geiger at the global scale, as the method uses continuous indices specifically targeting aspects of the global hydroclimate relevant to streamflow generation. This means that for the purposes of flood frequency analysis, it might better help to select groups of catchments for pooling. This proposition is considered in chapter 7.

2.6. Postscript

This covers the end of the literature review. The next section focusses specifically on the details of the GFM being used in this thesis, followed by a data and methods section. A separate chapter is dedicated to the Bristol GFM, as a clear understanding of it will be required to follow the significance of the work undertaken.

3. Modelling Framework

The transition from local to global scale modelling represents a step-change in the capabilities of flood modellers. While there have been significant studies in local scale inundation modelling, the move to continental and global-scale models produces additional considerations in the type of data used and the scale of modelling undertaken. As previously mentioned, global scale modelling methodologies can be defined as “from data to hazard”. This means that everything necessary for the prediction of flood hazard is integrated into a singular modelling chain. This paradigm of globally available and globally consistent flood predictions demands a flexible methodology, with novel parameterisations which accommodate coarser data that are generally less reliable than that which is available at the locally gauged study basin.

In this chapter, the structure and some key parameterisations of the Bristol Global Flood Model (GFM) are discussed. Despite the fact that the original paper describing this GFM was released in 2015 (Sampson et al., 2015), a general description of the model structure is yet to be published, which given the novelty of the model’s structure needs to be rectified. In doing so, it will be clearer why some of the decisions in the following study were made, as it is arguable that should the same investigations have been made in the context of an already well understood inundation model, the study would be substantially less important with respect to the existing literature. However, it is the case that the discipline of global flood modelling is still in its infancy and there are still fundamental questions to be asked regarding the dominant model sensitivities and uncertainties given the novelty of the modelling framework.

3.1. Overall introduction and flowchart

Although each chapter in this study is concerned with the Bristol GFM, because all of the GFMs abide by the general rule of “from data to hazard”, the findings of these studies will have strong implications for the other models, as they too will be attempting to model many of the same processes using coarse, remotely sensed data, and are also dealing with the decomposition of larger domains due to the limits of computation. In figure 3.1, a high-level flow chart of the model is presented. This highlights that the modelling of inundation is not contained in just the underlying hydrodynamic model. Rather, there are various pre-processing elements which manipulate the remotely-sensed data, which can be thought of as

a part of the modelling framework. Indeed, the framework of the Bristol GFM can be represented with four separate components. These are the pre-processor, the regionalised flood frequency analysis, the hydrodynamic model, and the post-processor. The pre-processor, regionalised flood frequency analysis, and post-processor are written up as functions in MATLAB, while the hydrodynamic model LISFLOOD-FP is coded in C++ for computational efficiency and can be called from a Windows or Linux operating system. It should be noted here that the Bristol GFM version of LISFLOOD-FP is a separate version from the locally implemented model. We have used a compiled executable from a clone of LISFLOOD-FP developed by Fathom, which is slightly modified to accommodate the GFM structure. In practice, this makes little physical difference to the results. However, it is different and any subsequent mention of LISFLOOD-FP in the context of the running of the Bristol GFM will be concerning this version. In the following sub-sections, descriptions of each of these functions and their necessary inputs and outputs will be described, along with their corresponding parameterisations.

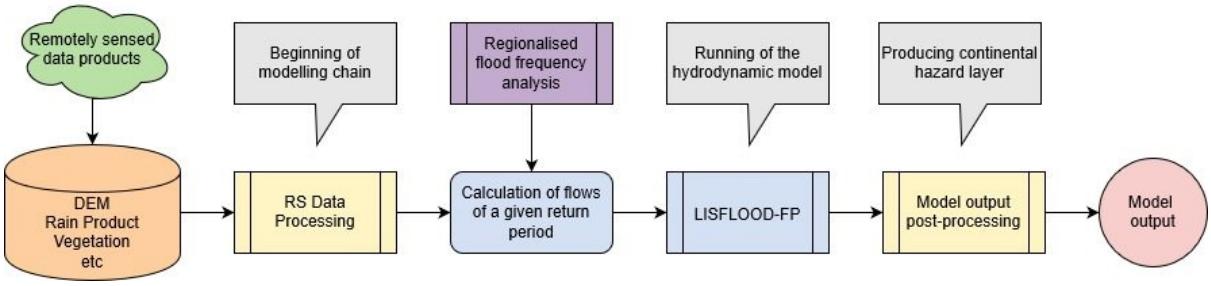


Figure 3. 1 - High level representation of the main elements which constitute the Bristol GFM.

3.2. The pre-processing elements of the model

The first stage of the process is the extraction and clipping of the remotely sensed data files at the modelled location. Many of the global datasets have been extracted in advance, meaning that there is a full global database which accompanies the model when it is run. To clip the datasets, the modeller must provide the coordinates of the bottom left and top right corner of domain in question. This rectangular shape of domain is currently the only way to produce inundation estimates in the methodology. If a singular catchment is desired, then a mask has to be used after the modelling chain has executed. The coordinates are given by the modeller in degrees, such that the latitude and longitude of the lower left and top right corner are given. For example, modelling of the lower Thames river would involve retrieving the necessary data using the coordinates [51, -1, 52, 0]. This information is taken by the input file

builder function, which cuts out the remotely sensed data sets needed by the GFM at this location from a repository of data with global coverage. This step is necessary as the remotely sensed data are normally stored in rasters of different size, and the coordinates given to the input file builder can often mean that the data requires reformatting (e.g. Clipping, mosaicking and aggregation to common resolution) for the specific location.

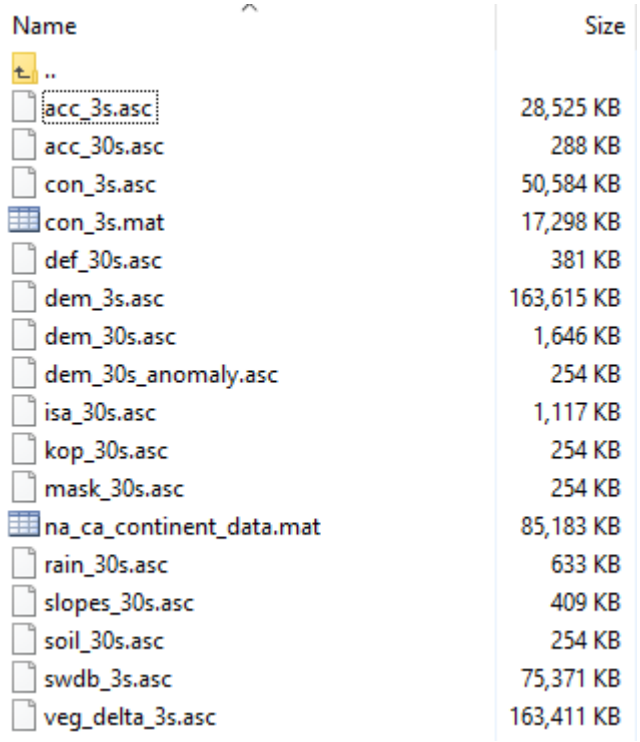


Figure 3. 2 - Input files generated from the remotely sensed data for a given location.

The result of running this function is a folder of raster files for the specific location modelled, which are used by the remaining functions of the pre-processor, to produce the necessary inputs files for the running of LISFLOOD-FP. An example of these files can be seen in figure 3.2, and are described table 3.2 below.

Filename	Description	Source
Acc_3s.asc	Upstream area accumulation data at 3s	Hydrographic data
Acc_30s.asc	As above but 30s resolution.	Hydrographic data
Con_3s.asc	Hydraulically conditioned DEM at 3s resolution	Hydrographic data
Con_3s.mat	As above in alternative format.	Hydrographic data
Def_30s.asc	Defence data estimated by the model	Model
Dem_3s.asc	Elevation data at 3s resolution	DEM
Dem_30s.asc	As above but 30s resolution.	DEM
Dem_30s_anomaly.asc	DEM features considered artefacts and therefore false by model algorithms	SRTM/Model
Isa_30s.asc	Proxy data for urbanisation given by impervious surface area	Urban data
Kop_30s.asc	Climate dataset using Koppen-Geiger classification	Climate data
Mask_30s.asc	Country mask data for created national hazard maps	Country data
Na_Ca_continent_data.mat	Relationships of catchment area to stream width.	Model
Rain_30s.asc	Precipitation dataset	Precipitation data
Slopes_30s.asc	Slopes estimates of channels derived from DEM	SRTM/Model
Swdb_3s.asc	Database of surface water to account for lakes and reservoirs	Hydrography data
Veg_delta_3s.asc	Estimation of DEM elevation reduction due to vegetation	Vegetation data

Table 3. 2 - The files needed for running of the Bristol GFM, derived for the study location. The data comes in a mixture of 3s and 30s resolution.

Pre-processing functions

While pre-processing is more commonly associated with cleaning of the data and not specifically a modelling process, because of the nature of the data used it must fundamentally be considered a modelling process in global flood modelling. That's because the model is aiming to recreate natural processes and the ability to do this skilfully is a function of how the natural environment is represented. Therefore, as the raw remotely sensed data is modified and made readable to the underlying hydrodynamic model, various assumptions are made, which leads to the large list of parameterisations to be covered later. As such, it is important to understand what modifications are made to the raw data to fully understand the breadth of the model's capability. In this sub-section, each of the functions used in the pre-processor are briefly described, with much greater descriptions provided where there is a significant process or parameterisation occurring. In figure 3.3, a flowchart is shown, which shows each of the functions described, as well as the principal inputs and outputs of each function, finishing with the running of LISFLOOD-FP and the subsequent post-processing.

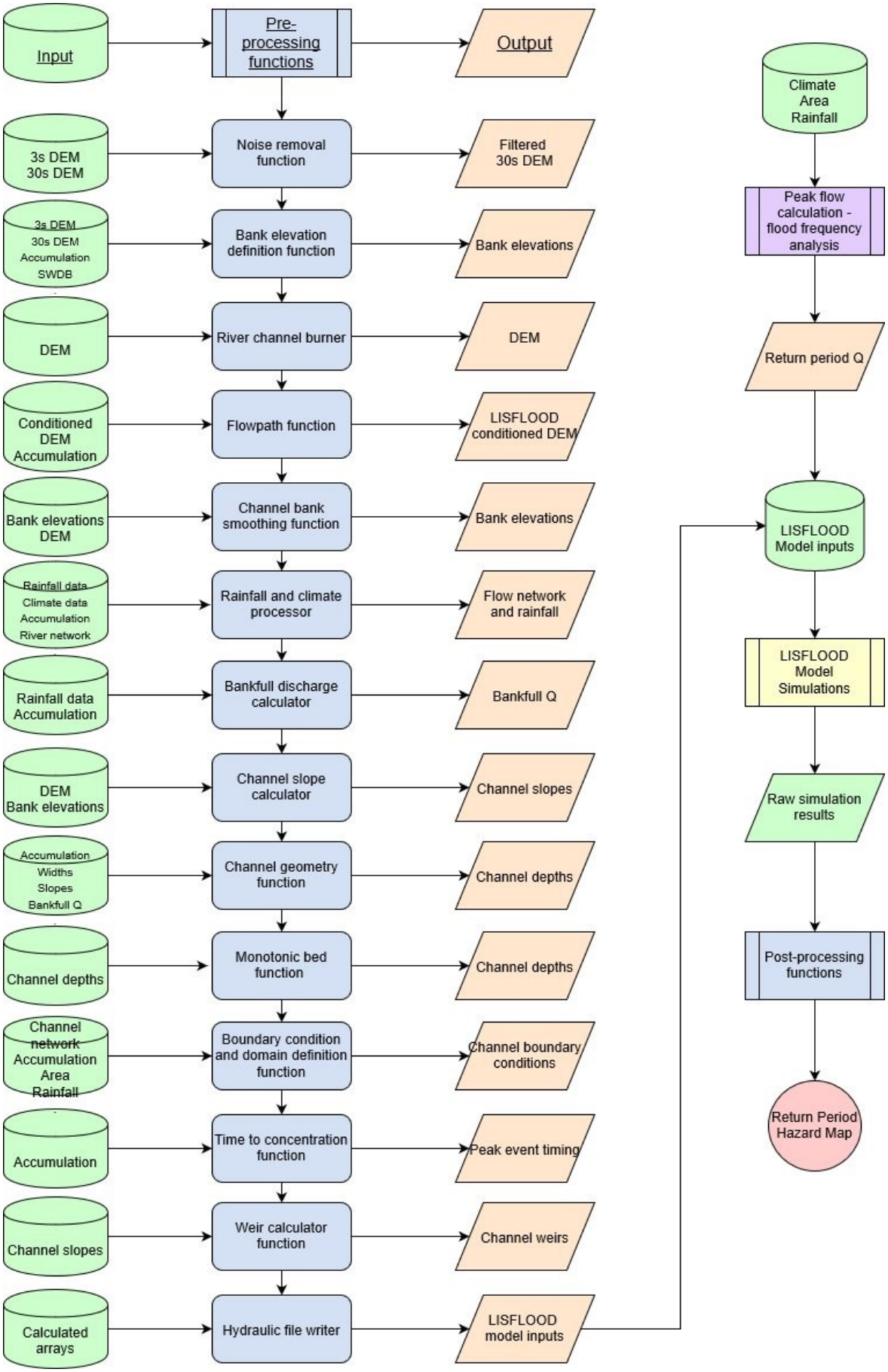


Figure 3. 3 - Flow diagram of the principal components of the model structure.

Pre-processing master function

Each of the functions for the pre-processor is called from a master function. There are two stages to the master function; the first stage focusses more on DEM manipulation and formatting tasks. The second stage focusses on process-based elements of the pre-processing, which cover the behaviour of the model, such as the explicit definition of the channel on the domain.

- Noise removal function

This function contains a basic filter which reduces urban elevation in the 30s DEM to the same extent as is done on 3s DEM. This function also contains an algorithm which attempts to remove noise in the SRTM-derived DEM without erasing actual features.

- Bank elevation definition function

The purpose of this function is to make sure that the bank elevations are well defined at 30s resolution, which is done by using the finer resolution 3s DEM which SRTM also produced. In the finer 3s DEM, the bank elevations are better defined, than in the 30s DEM which is used in this study. This means that transferring the bank elevations from the 3s to the 30s DEM will ensure that the bank heights are better represented. Whether or not individual cells are modified in this way is based on the accumulation size at that point. This function also ensures that bank heights next to or under surface water bodies such as lakes and reservoirs are ignored, as these cell heights shouldn't be used to modify the river banks.

- River channel burner

This function modifies the DEM so that it represents larger rivers. At 30s DEM resolution, the majority of rivers and streams are not represented and require a sub-grid scale method (Trigg et al., 2012). However, there are some rivers of a much larger size. In the DEM, these are not accurately represented as only the water surface. Therefore, the actual channel needs to be "burned" into the DEM, so that hydrodynamically speaking, the water flows along these channels rather than simply being parameterised at the sub-grid scale.

- Flowpath function

This function changes the flowpaths which are given by the hydrographic data in the HydroSHEDS dataset to work with the structure of the hydrodynamic model. In the topographical processing of the HydroSHEDS dataset, the flowpaths of channels have 8 directions, horizontal, vertical and diagonal (Lehner et al., 2008). However, the raster modelling approach used by LISFLOOD-FP means that this flow routing has to be reduced to 4 directions to work effectively, which has a subsequent effect on the accumulation.

- Channel bank smoothing function

As the banks heights are defined by either a noisy DEM or selected from a prior function, it is possible that the bank heights will be quite noisy as well. In terms of realism, it was decided to have a smoothed bank, which better represents channel bed topography. Otherwise, it could be that there are random cells at the banks where flooding doesn't occur due to deep channels.

This processing is governed by some smoothing parameters, which idealise the bank heights as a polynomial. When the in-situ bank heights diverge dramatically from this polynomial function, they are modified to make them smoother.

- Rainfall and climate processor

This processing element has two functions. The first is to ensure that the representation of upstream rainfall passing through each river cell accounts for all rain in a catchment including if the catchment extends outside of the hydrodynamic model domain. Therefore, the precipitation associated with a given river cell is the average precipitation of the upstream catchment area. This means that when it comes to estimating the peak flow at a given location, it accounts for the general amount of precipitation in the catchment instead of at that particular location.

The second function reduces the network density of the model in arid regions. Because a channel is defined in the methodology as a cell with an accumulation level above a given value, there is a tendency for channels to be produced in arid regions which don't exist, because the topographic processing means that cells can have a high accumulation value, regardless of climate. In these cases, the average rainfall is used as a filter to determine

whether or not there should be a channel at the given location. If the rainfall is below a given value, then the channel is removed from the domain as it likely doesn't exist.

- Bankfull discharge calculating function

This function assigns a bankfull discharge value to each cell which has a channel. The bankfull discharge is calculated through the use of upstream accumulation and rainfall. With these, the bankfull discharge is calculated through the regionalised flood frequency analysis (rFFA) (described section 3.3). As the main function of this is to produce the return period flows, by definition it also produces bankfull flows when bankfull is determined by a particular return period discharge. This is done through the estimation of a bankfull discharge return period, with the most common value being between 1.5 and 2 for temperate climates (Williams, 1978). Hence the return period of bankfull discharge is a parameterised value in the model. A fuller description of this method can be found in the rFFA section on this module.

- Channel slope calculator function

This function is used to estimate channel cell slopes. This is done by taking the derivates of 'n' upstream and downstream cells and their corresponding bank elevations. The function also has a slope threshold which defines the upper limit of slope permissible in the DEM, which should help reduce the impact of artefacts not yet identified by other elevation filters.

- Channel geometry function

This function calculates the depth of the river channel cells, given the width, slope, bankfull discharge and channel roughness of the corresponding cell. This function uses another of the principal parameterisations of the model, that the depth can be calculated by inverting the Manning's equation and making the depth subject. Consider the Manning's equation below:

$$Q = A \cdot n^{-1} \cdot R^{\frac{2}{3}} \cdot S^{\frac{1}{2}}$$

Where Q is the discharge in (m^3/s), A is the area in (m^2), n is the dimensionless Manning's roughness coefficient, R is the hydraulic radius (m), and S is the dimensionless channel slope. Manning's n is a parameter which is estimated by the user, with a default value of 0.03. The Q value has already been calculated in a previous function for each cell by the regional flood frequency analysis, and so has the slope. To calculate the area and the hydraulic radius, a

strong assumption is made about the channel – that it is rectangular, and the friction effects of the channel sides are negligible. This means that the area and the hydraulic radius can be represented like so:

$$A = w.d$$

$$R = \frac{A}{P} = \frac{w.d}{w} = d$$

Where *w* is the channel width and *d* is the channel depth. This means that the Manning's equation can be written like so:

$$Q = w \cdot d \cdot n^{-1} \cdot d^{\frac{2}{3}} \cdot s^{\frac{1}{2}}$$

If the depth variable is re-arranged to become the subject of the equations, it can be represented as:

$$d^{\frac{5}{3}} = \frac{Q \cdot n}{w \cdot s^{\frac{1}{2}}}$$

Rearranging the power terms, the depth at each location is calculated with the following equation:

$$d = \left(\frac{Q \cdot n}{w \cdot s^{\frac{1}{2}}}\right)^{\frac{3}{5}}$$

Therefore, from this calculation at every channel cell, the depth of the channel is found. It is worth noting however, that the width is estimated, rather than being precisely known. Despite the recent development of fairly comprehensive width databases (Frasson et al., 2019), in this version of the model, the widths are calculated from continental-scale splines which relate the channel width to the upstream area. Therefore, the width is itself also a function of upstream accumulation.

- Monotonic bed function

This function ensures that the bed is constrained to a monotonic structure. This means that while it is possible for the slope to change, it is not able to become a positive slope moving downstream. Because of the calculation of the channel beds from bank heights, it is not always the case that the channel bed decreases downhill. There are points when the channel

can stop being monotonic and the channel move on an upwards gradient. For the purposes of channel and inundation modelling, this is problematic and it is important to ensure that the slope is always in the direction of the flow path, so that the river can be most effectively simulated.

- Boundary condition and domain definition function

The previous functions of the pre-processing have involved conditioning of the data. The purpose of this function is to calculate the boundary conditions for the model required at each of the tributaries and at various lengths along the main rivers of the domain. The reasoning behind this method of producing boundary conditions is predicated on the fact that the underlying hydrodynamic model is LISFLOOD-FP, which has structural aspects that can be obstacles to modelling at the global scale. Each of these is to a certain extent accounted for in this function. The first of these limitations is that as a hydrodynamic model, increasing the Q along the lengths of the reach becomes difficult, as it is impossible to know a priori the timing of the flood wave as it propagates downstream. This is accounted for by the decomposition of the domain into smaller sub-reaches, each individually modelled. This means the effects of an increasing river flow due to infiltration and river length (which in a global model could be several hundred miles) is negated. As a result, there has to be regular additional boundary conditions along the river reach to ensure that the flows at these points correspond to the size of the river at that location.

To clarify, the location of boundary conditions within the model has to be declared. This is because the inundation method does not permit additional flow to accumulate as the flood wave travels downstream. Therefore, one requires a new release wave for larger accumulations of flow. How often a new boundary condition is set is parameterised in the model via the reach length. There are first parameters to constrain the minimum and maximum reach lengths between the boundary conditions. There is a clause which determines how small the reach can be, if the channel is determined to be a “stream”, when it is below a certain accumulation, then the reach length can be even shorter. Besides this, there is a further parameter called the “reach accumulation threshold”. This is designed to allow a varying reach length between the minimum and the maximum with respect to the

change in accumulation. Starting from the downstream (DS) boundary conditions, as one moves upstream (US) there is a change in accumulation of the reach. The logic of the parameter is that it creates a new reach when the ratio of US to DS points is met. Therefore, a reach accumulation threshold of 0.8 means that that a new US boundary is created when the US accumulation is 80% of the previously defined DS accumulation. This process has been very slowly understood by analysis of the model, without explanation from the model developers. As documentation was not available, it had to be "re-discovered" through modelling and analysis. This is why it will be seen that in earlier use of the parameter, it involved using values above 1, when in fact the highest value which the parameter could take would be 1, besides which the parameter reverts to the constraining reach lengths.

To re-iterate the core purpose of the function, it is to determine the points at which the boundary conditions are generated. These boundary conditions generate return period discharges using the regionalised flood frequency analysis. This means the domain can be split up and run as separate models, and it means that the peak discharge is able to change along the river system. When a tributary is encountered, the distances to the next upstream boundary condition are found by using the proportion of accumulation that comes from the given tributary. This assumes that the size of a tributary's catchment is directly proportional to its contribution to the larger river.

- Time to concentration function

While the previous function determines the location of the peak flow inputs, and the peak flow, the shape of the hydrograph also needs to be estimated, as the shape of the hydrograph is instrumental in determining the hydrodynamics of the event. In this model, the timings of the hydrograph are proportional to the time to concentration (Grimaldi et al., 2012). Once the hydrograph timings are known, the hydrograph is constructed as an idealised triangle, with the receding limb being 4 times longer than the rising limb (Chow et al., 1988). The time to concentration is the time taken for water to travel from the uppermost part of the catchment to the downstream point under consideration. This is calculated by accounting for the distance between these points and also the topographic effects – i.e. whether the catchment is small and peaky or large and attenuated. This estimate is then used to calculate the rising limb length.

- Weir calculator function

A key source of instability in the model is when conditions are such that the flows being modelled move from subcritical to supercritical flow. While in local studies this can be accounted for, in a global scale consistent methodology it presents a considerable problem. This problem is alleviated by modelling the supercritical flow with a weir equation. In the modelled domains, the super-critical flow will occur when the slope exceeds a certain point. For these cells, rather than modelling the flow as super-critical, the flow is controlled by a standard weir equation. While certainly false in the sense that there are no weirs, and it is simply a workaround to ensure supercritical flows do not create instabilities in the running of the model, it offers a solution to otherwise unrealistic and non-behavioural simulations quite elegantly. It is worth noting that later versions of the Bristol GFM have replaced this function with a Froude number limiter; however, for the duration of the analyses undertaken here, the weir function is used.

- Hydraulic file writer

The end result of the preceding functions is that there are a number of arrays holding information about the domain to be modelled. There is spatially correlated information regarding bankfull discharge, channel geometry, boundary conditions and other information. The purpose of this function is to format all of this information into files which are then run on the underlying hydrodynamic model, which is the compiled global version of LISFLOOD-FP.

This concludes the various functions of the pre-processing element of the model. As can be seen, there are multiple parameterisations which are novel in the context of inundation modelling, such as the calculation of depth from Manning's, but concern issues which are endemic to global flood modelling.

3.3. Calculation of peak flows

In the previous section, the locations and the timing of the peak flows for a given return period were derived. The next step is to derive the magnitude of the flow. For the Bristol GFM, the chosen method is a regional flood frequency analysis (rFFA) at the global scale (Smith et al., 2015). This method is a development of other prior regional flood frequency analyses, but which had smaller scale study areas (Farquharson et al., 1992; Meigh, 1995). These studies

are founded on the index-flood method (Kjeldsen & Jones, 2007), which finds a flood magnitude of a given return period with the following constituents:

$$Q_{RP} = Q_{MAF} \times GF_{RP}$$

Where Q_{RP} is the discharge of a given return period; Q_{MAF} is the discharge of the mean annual flood, which is the average peak flow that is encountered each year; and GF_{RP} is the growth factor for a given return period. Hence, to implement the global scale rFFA, Q_{MAF} and GF_{RP} are needed at every boundary condition location, defined by the boundary condition function. These are calculated using the method of Smith et al., 2015, which is briefly described in the following sub-section, and again in detail in chapter 7.

To regionalise the flood frequency analysis across the globe, three variables are used. These are the climate zone as defined by the Koppen-Geiger classification; upstream catchment area of the point of interest; and average area rainfall, accounting for upstream rainfall as described in the rainfall and climate processor. Historical river flow data is the other pre-requisite for regionalisation, taken from gauged stations which possess similarity to the ungauged basins of interest. Gauge data were obtained from the Global Runoff Data Centre (GRDC, 2018), which compiles streamflow and catchment information from around the globe.

- Calculation of Q_{MAF}

Q_{MAF} is important to calculate because it determines the magnitude of the flow, while growth curves determine catchment response. While the response of two catchments for return period flows could be similar, if the Q_{MAF} is different the magnitude of the return period flow will be different. This will be reflected in the Q_{MAF}. In the method of Smith et al., 2015, there are Q_{MAF} surfaces generated for each Koppen-Geiger zone, which use the catchment characteristics of area and average annual rainfall as predictors. Although surfaces are generated for each climate zone, they all have a similar shape, which will be explained in chapter 7. Figure 3.4 shows the Q_{MAF} surface for the temperate climate, classified using Koppen-Geiger. The Q_{MAF} surfaces of other climate classifications are similar in form, in that the value of Q_{MAF} changes positively as average annual rainfall and catchment area increase. This is a part of the design of the surface, which ensures a positive relationship between these variables.

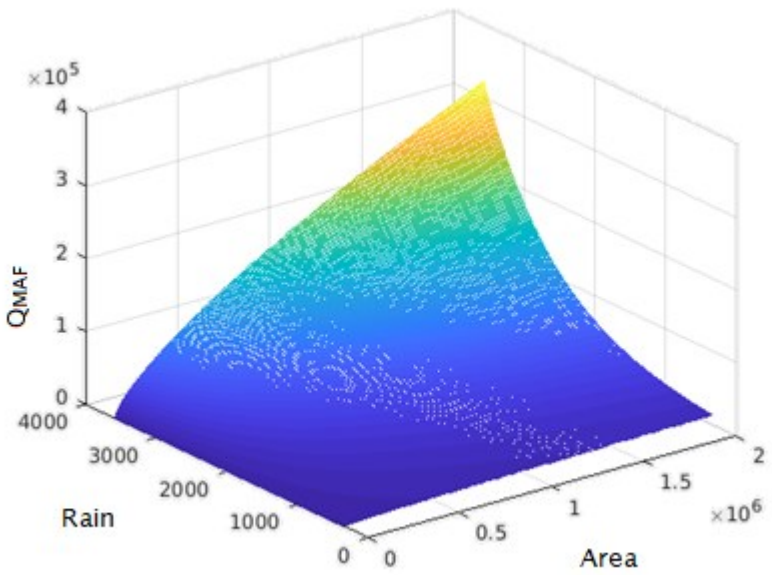


Figure 3. 4 - MAF surface of the temperate climate (according to KP classification).

Chapter 7 will investigate the underlying methods of the rFFA in greater detail, therefore the assumptions and processes which lead to the production of this MAF surface will only be covered in brief here, taking the temperate climate surface above as an example. Approximately half of the 5,500 stations in the GRDC database are classed as temperate. Each of these stations has an upstream area, average annual rain and Q_{MAF} value which can be associated with it. The first fit used is a power curve, which fits area to Q_{MAF}. This is done in groups of the data which is binned by the precipitation magnitude. While these curves produce a surface which increases with area, it doesn't always increase with precipitation. Therefore, the surface needs smoothing. To do this, a linear fit is used across bands of data according to the area. In this fit, the natural log of Q_{MAF} is fitted as linearly proportional to the precipitation curves, calculated from the power fit. This method is more closely examined in chapter 7.

- Calculation of growth factor

While the Q_{MAF} determines the magnitude of the average annual maximum flow, the growth factor determines how much the total discharge increases from this with larger return periods. This is in effect the result of catchment response. In smaller, topographically varying catchments, it is usual to find that the growth curves are much steeper, than in very large,

low sloping landscapes. The steepness of the growth curve is often referred to as the catchment's "peakiness" (Hrachowitz et al., 2013). These effects are found by observing the annual maximum (AMAX) time series of the stations. However, in the GRDC, many of the stations have time series which are too short for a reliable calculation of the growth curve, and considering individual gauges doesn't address the growth curves of ungauged catchments. To regionalise the growth curves, the catchments of each climate zone are clustered, using log(precipitation) and log(area) as variables. This creates multiple clusters of the stations in each climate zone. To create the growth curves, their AMAX time series are normalised, and then joined into a continuous time series. In some cases, each cluster can have upwards of 4000 AMAX datapoints. A generalised extreme value (GEV) distribution is then fitted to these data, so that the growth factors are made to correspond to a given return interval. To calculate the growth factor at an ungauged basin, its upstream area and average annual rain are used to measure its proximity to the centroids of the calculated clusters. The final growth factor value is a distance-weighted sum of the growth factors of every cluster. This is explained in greater detail in chapter 7.

- Synthesis of Q_{MAF} and growth factor

Using the two methods above, the Q_{MAF} and growth factor are estimated for any location, which has an average annual rainfall and upstream area. In the model, each of these is given by the data compiled at that location. To re-iterate, the upstream accumulation of that cell is used for area. For average annual rainfall, the mean rainfall of all the cells upstream is used. From these, the MAF surfaces and corresponding growth curves are called and used to make the predictions of Q_{RP} at every boundary condition in the model. It should be noted here that the value of Q_{RP} is applicable to both the calculation of bankfull discharge as well as the event magnitudes which are used to generate the hazard maps from model outputs.

3.4. Hydrodynamic simulation

As already stated, the hydrodynamic inundation model is a global version of the LISFLOOD-FP algorithm (Bates et al., 2010; Sampson et al., 2015). At the global scale, this means that the model is running at a resolution of $1km^2$ natively. As a coupled 1D-2D hydrodynamic model, this presents some difficulty, as the features of many channels below a certain size are lost. Rivers of a width lower than 1km cannot be physically represented in the DEM. Instead, they need to be defined as attributes of each cell in the gridded domain. This representation is

termed as a subgrid parameterisation (Neal et al., 2012). Although the term subgrid parameterisation is used, this is still in essence a coupled 1D-2D hydrodynamic scheme, but the key difference is that the subgrid channels do not have a spatial component to them. They are instead calculated within the domain which would be defined as the 2D inundation domain, which is the reason for the scheme's designation as "subgrid".

Actual execution of the inundation modelling is performed once all the preceding information has been compiled. The modelling is performed using the globally compiled version of LISFLOOD-FP. Using the files written by the pre-processing function as inputs, LISFLOOD-FP is called individually for each decomposed sub-reach. The running of the decomposed domains occurs on the University of Bristol servers, and involves distribution of the model simulations across Condor, a high throughput computing resource. Individual domains are modelled, and the resulting output of interest is a raster of the maximum flooding depths. Because the application of this model is specifically for flood hazard, this is appropriate. Unlike in the case of forecasting where the initial conditions are better known (Alfieri et al., 2013), the estimation of hazard implies the hazard of a given return period, which is itself a synthetic event. Therefore, the primary interest is whether a cell was flooded and to which elevation and depth it was flooded, regardless of the inundation time or duration of inundation.

3.5. Post-processed outputs

As already described, the model must decompose the domain into individual components for LISFLOOD-FP to run. Thus, it is left to the post-processing component to stitch together these decomposed domains and present a map of return period maximum depth (hazard). There will be overlap in the flooding hazard between one decomposed domain and another: in some capacity it can be anticipated that there will be flooding further downstream than the next boundary condition. However, in such cases, the higher flooding from whichever subdomain will be selected at that point, creating an overall aggregating maximum depth map for a given return period.

3.6. Model's parameterisation diagram

In the previous sections, the main processes of the GFM and the corresponding assumptions have been summarised. In places, some of the critical parameterisations have been explained in more detail, as their significance on model behaviour is anticipated to be quite high. However, alongside these are many other minor parameters which are not easily explained,

but have the potential to influence model behaviour. It is for this reason that it is necessary to perform some basic analysis on the GFMs, to assess which parameters are important. As such, table 3.3 below presents all the parameters identified in the Bristol GFM, which will all be analysed in chapter 5 to assess their relative sensitivities. These parameters are distributed throughout the functions previously described. It is important to understand their role in the model structure, as they are central to the consistent modelling at the global scale, within the "from data to hazard" paradigm.

Parameter	Default value (units)	Parameter description	Position in structure
urban def par a	0.05 (n/a)	Change size of channel at bankfull discharge due to urbanisation	Urban
urban def par b	0.65 (n/a)		Urban
Q_{bf}	2 (years)	Number of years for bankfull discharge to occur	Hydraulic
accumulation threshold	50 (US cells)	Minimum accumulation to define a river in model	Hydraulic/ topographic
slope threshold	0.03 (m/m)	Slope threshold to remove DEM artefacts	Topography
surface water threshold	100,000 (US cells)	Determines if a water body is a river or other	Hydraulic
neighbour filter	10,000 (US cells)	Used to derive bank heights	Topography
bank smooth par a	0.1 (n/a)	Smoothens bank heights along river profile – maintains monotonic rivers	Topography
bank smooth par b	0.38 (n/a)		Topography
bank smooth par c	5 (n/a)		Topography

arid screen par - rain	300 (mm)	Removal of channels in the case of arid environments	Climate
arid screen par – acc	1,000 (US cells)		Climate
hyper-arid screen – rain	40 (mm)	Removal of channels in the case of hyper-arid environments	Climate
hyper-arid screen - acc	50,000 (US cells)		Climate
manning's n	0.035 ($m^{1/3}/s$)	Determines channel roughness	Hydraulic & Topographic
min reach stream length	10 (km)	Minimum stream size defined for model	Hydraulic/ topographic
max reach stream length	20 (km)	Minimum river size defined for model	Hydraulic/ topographic
river/stream threshold	1,000 (US cells)	Determines accumulation for stream or river	Hydraulic/ topographic
reach length max	200 (km)	Maximum reach size within the model	Hydraulic/ topographic
reach acc threshold	0.8 (n/a)	Threshold determining accumulation increment per each river length	Hydraulic/ topographic
area/rainfall weight	0.5 (n/a)	Weighting of rainfall vs. area in regionalisation	Flood frequency analysis
min stream velocity	1 (m/s)	Stream velocity minimum limit	Hydraulic
meander coefficient	0.7 (n/a)	Coefficient of channel meander at the subgrid-scale	Topography
floodplain n	0.06 ($m^{1/3}/s$)	Determines floodplain roughness	Hydraulic
veg filter par a	0.1145 (n/a)	Determines reduction of DEM height in presence of vegetation	Vegetation
veg filter par b	0.4335 (n/a)		Vegetation

urban extent low	0 (%)	Degree of urbanisation whereby cells are removed, assumed to be buildings	Urban
urban extent med	10 (%)		Urban
urban extent high	50 (%)		Urban
ground truth low	0.2 (quantile)	Which values are used to find ground truth	Urban
ground truth high	0.5 (quantile)		Urban
flood cluster parameter	5 (n/a)	Determines how many clusters formed for regionalisation	Flood frequency analysis
min AMAX record length	30 (years)	Minimum AMAX record length for consideration in regionalisation	Flood frequency analysis
MAF error threshold	0.25 (n/a)	Max allowable error in MAF calculation	Flood frequency analysis
rain outlier lower	0.9 (n/a)	Determines what rainfall measures are determined to be outliers	Flood frequency analysis
rain outlier upper	1.5 (n/a)		Flood frequency analysis

Table 3. 3 - Table of all parameters including their default values, descriptions and approximate position in model structure.

4. Data and methods

Having presented the complete model structure and parameterisation for the first time, we can now discuss the various datasets necessary for running of the model. Alongside this is presented the data and methods used for model analysis, for the purposes of validation and conditioning. Following this is a description of the various methods which are used in the results chapters, along with the reason for their choice and implementation.

4.1. Data

4.1.1. Remotely sensed data for modelling domains

As discussed, the pre-processing component of the model is where most of the datasets are incorporated. The ability of the model to be considered global is centred around the global extent of the datasets used. In table 4.1 below, each of these datasets is described in brief. Note that the resolution is in radial coordinates, meaning the resolution shown below is in minutes and seconds.

Dataset	Data type	Resolution	Description	Reference
SRTM	DEM	3”	Global elevation dataset widely used	(Rodríguez et al., 2006)
HydroSHEDS	Conditioned DEM Accumulation	30”	Conditioned SRTM for hydrographic use. Topographic analysis derives upstream accumulation	(Lehner et al., 2008)
CMORPH v1	Average Annual Rainfall	2.5’	Rainfall dataset derived from microwave sensors	(Joyce et al., 2004)
Koppen-Geiger	Climate type	15’	Updated KP climate classification scheme, which is correlated to temperature and precipitation	(Peel et al., 2006)
GRDC	River station data	-	Annual maxima data compiled by the GRDC,	(GRDC, 2018)

			with ~5500 stations extracted	
MODIS VCF	Vegetation cover	7.5”	Dataset measuring percentage tree cover per cell	(Di Miceli et al., 2011)
USGS ISA	Urban land cover (concrete)	30”	Percentage of impervious surface area per cell, derived from luminosity data	(Elvidge et al., 2007)

Table 4. 1 - Table of all the datasets which are necessary for the model to make predictions of flood hazard at a location.

4.1.1.1. HydroSHEDS conditioned terrain data

The most important dataset required for hydrodynamic inundation modelling is the DEM, as this determines the domain over which the rivers and floods will flow. To perform inundation at the global scale requires global-scale data, which the SRTM comes the closest to representing; the dataset is constrained by the latitudes of 56°N to 60°S. Nonetheless, between these latitudes the Earth’s elevations are fully covered. The quality of the data has been closely examined (Rodríguez et al., 2006), and in its raw form it is not suitable for the representation of hydraulic processes and inundation modelling. Instead, a processed form of SRTM is the product which is instead used by the model, that of HydroSHEDS (Lehner et al., 2008). While being very similar to the SRTM dataset, its purpose was to ensure that the various errors of the datasets found do not mean that the hydrography is fundamentally incorrect. This means ensuring that the river flow-paths are coherent, such that the upstream accumulation of each cell is correctly recorded, as the upstream accumulation is a fundamental attribute to the model. In figure 4.1, we can see the HydroSHEDS conditioned DEM across a 2° by 2° tile, representing a portion of the Sierra Nevada mountains in northern California. It is clear on the western side of the mountain where the flow-paths are. This is not yet the finished product used by the model, as there are still some modifications made to the raster, when the effects of vegetation and urbanisation are accounted for.

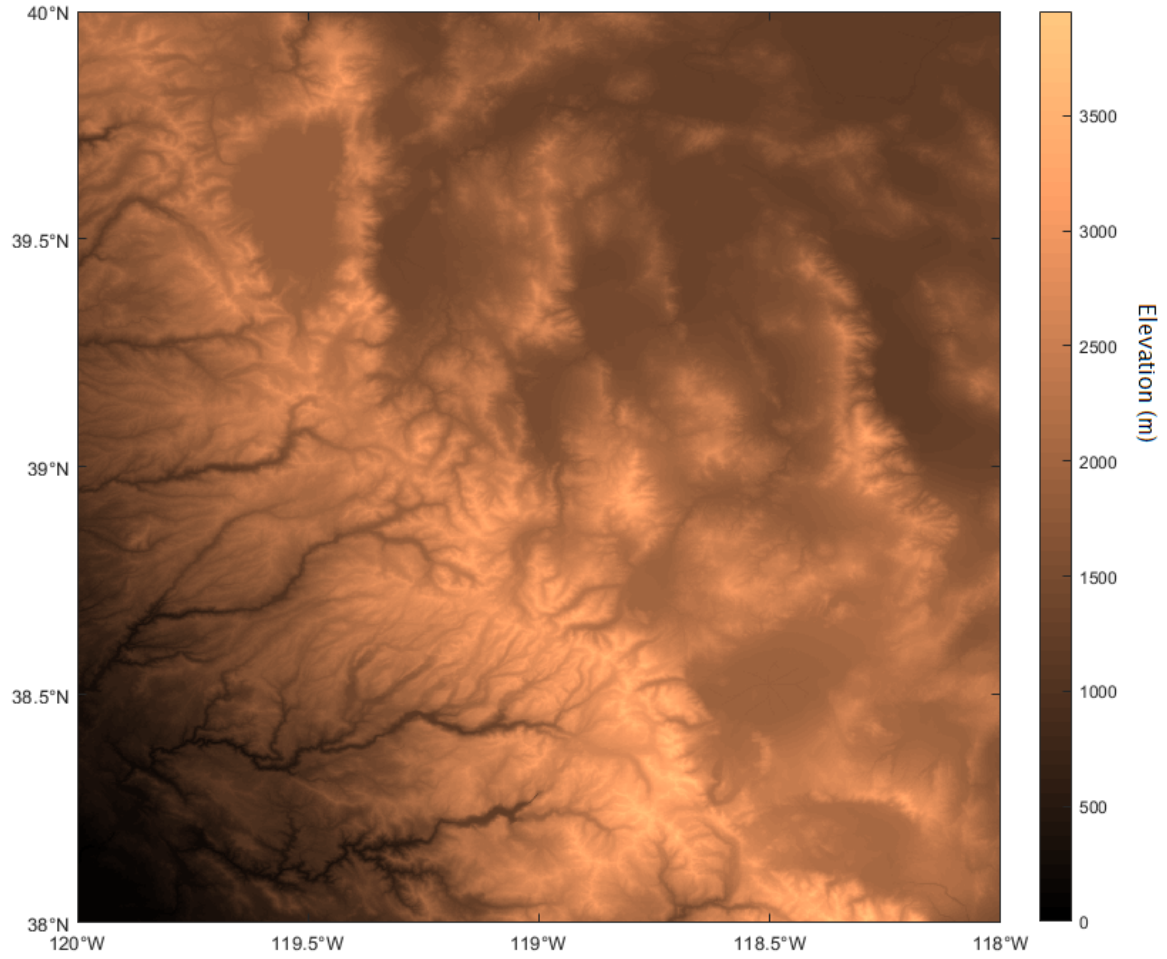


Figure 4. 1 - Hydrographically conditioned DEM from the HydroSHEDS data product.

4.1.1.2. HydroSHEDS derived hydrography

When investigating the hydrographically conditioned DEM from HydroSHEDS, it is possible to determine the upstream accumulation of each cell via topographic analysis. This is also recorded in the HydroSHEDS data and is shown in figure 4.2. The accumulation is an important dataset for the GFMs as it can be used to determine the size of river which is modelled, as well as the variable used for determining the upstream catchment area for the target rivers – which is used for regionalisation (described in chapter 3 as the upstream accumulation). In this instance, it can be seen that the accumulation rises very rapidly over a smaller area – this is due to the effects of convergent flow-paths in the Sierra Nevada.

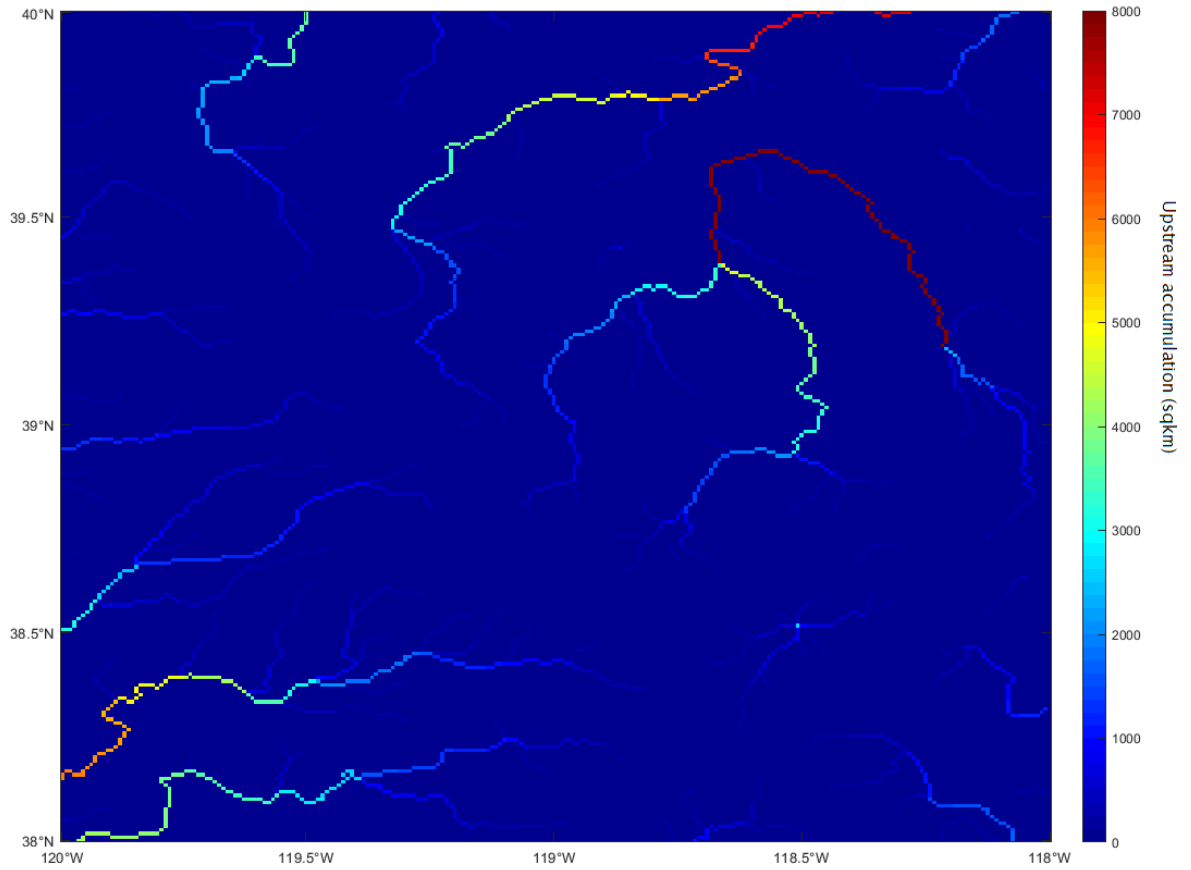


Figure 4. 2 - Hydrographically conditioned DEM from the HydroSHEDS data product, via TopoToolbox (Schwanghart & Kuhn, 2010)

4.1.1.3. Surface Water

The final dataset which is derived from the HydroSHEDS product is a database of water surfaces. These can be either reservoirs, lake or a part of a larger deltaic systems. These are used so that the lower flat areas which are lakes do not get mistaken for channels. In figure 4.3, there is the representation of several lakes including Lake Tahoe at the top of the Sierra Nevada mountains, as well as the smaller Walker's lake and Mono lake. Explicit modelling of these water bodies is also useful in that they normally have an attenuating effect on large-scale flooding (Bayliss et al., 2008) and so are important to include for better prediction of the hazard. However, given that the original data collection mission was over 20 years ago, there is potential error across many modelling domains, as some surface water bodies have changed significantly over a 30 year period (Pekel et al., 2016); for example the case of the drying Aral Sea (Micklin, 2007).

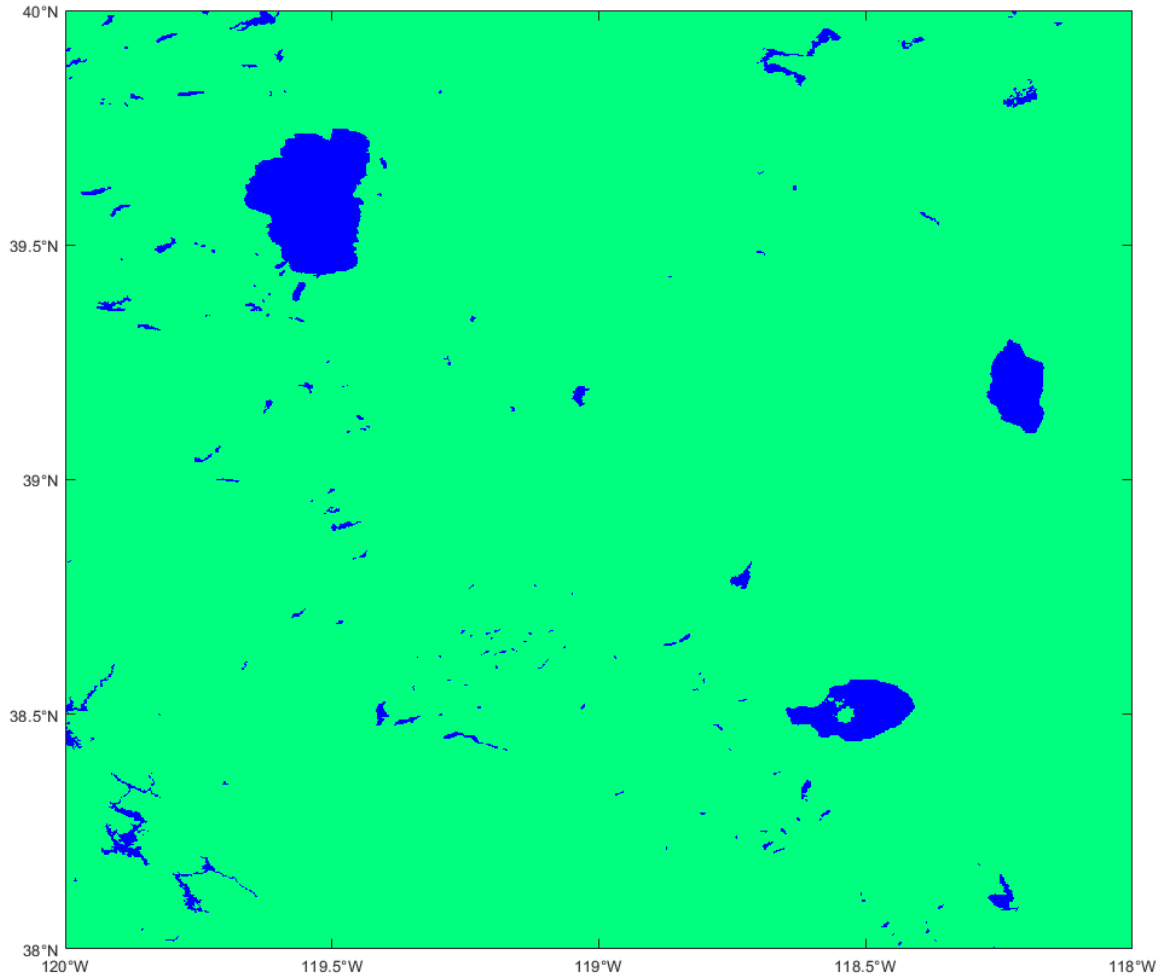


Figure 4. 3 - Water bodies represented in the Surface Water Database, showing lakes and reservoirs within the domain

4.1.1.4. Precipitation dataset

Concerning the representation of rainfall, there are multiple datasets which could be chosen. When examining 30 contemporary global precipitation datasets, it has been found that they can vary in their measure of precipitation by up to 300mm/year (Sun et al., 2018). In the case of the Bristol GFM, the rainfall measure used is an average annual estimate. This reduces the number of meteorological processes which can be represented by the data, such as storm intensity, an important process in the context of arid region flooding (Zaman et al., 2012). Seasonality is also poorly described by an average annual measure of precipitation (Berghuijs et al, 2016), yet it is nevertheless a highly useful indicator of flood prediction (Carey et al., 2010).

Figure 4.4 shows the average annual rainfall across the Sierra Nevada mountains, and the data which is seen by the model. The most intense rainfall occurs across a band of the domain,

which with reference to figure 4.1, can be seen as the effects of the mountain range, themselves a relatively short distance from the Pacific Ocean, meaning it can be concluded this pattern of precipitation is due to orographic effects.

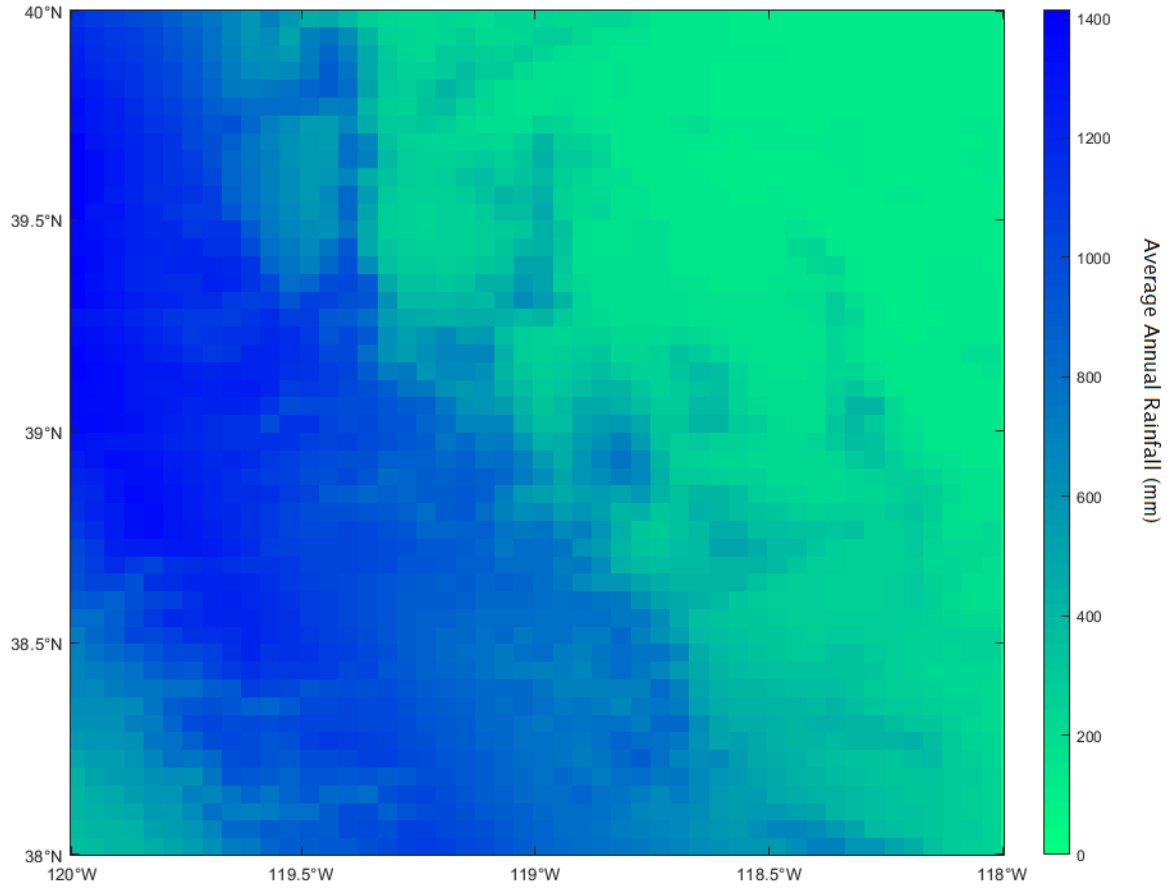


Figure 4. 4 - Average annual rainfall, according to CMORPH rainfall product.

4.1.1.5. Urbanisation dataset

While the correct representation of urban areas is highly important for quantifying exposure and vulnerability for modelling the risk, it also influences the hazard, and therefore the Bristol GFM as well. One of the flaws of the SRTM elevation data is that it is unable to differentiate ground truth from the tops of buildings or vegetation. Hence it is necessary to account for this. One way to do this is to parameterise the building heights, with respect to an indicator of urbanisation. This discrepancy is then removed from the DEM and the height of that raster is resampled from surrounding cells (Sampson et al., 2015). That is the first way that the ISA dataset is used. In figure 4.5, the more urbanised areas are shown as darker. Because the

domain is centred on the Sierra Nevada, there are not many urban sprawls, the only significantly urbanised areas correspond to Carson City and Reno on the upper border of the domain. The other use for the ISA dataset is to address the problem of flood defences. The parameterisations of the model, as discussed in chapter 3, try to incorporate likely flood defences into the model by assuming that most urban zones will have some level of flood defence standard, as the representation of flood defences is important for model realism (Ward et al., 2013).

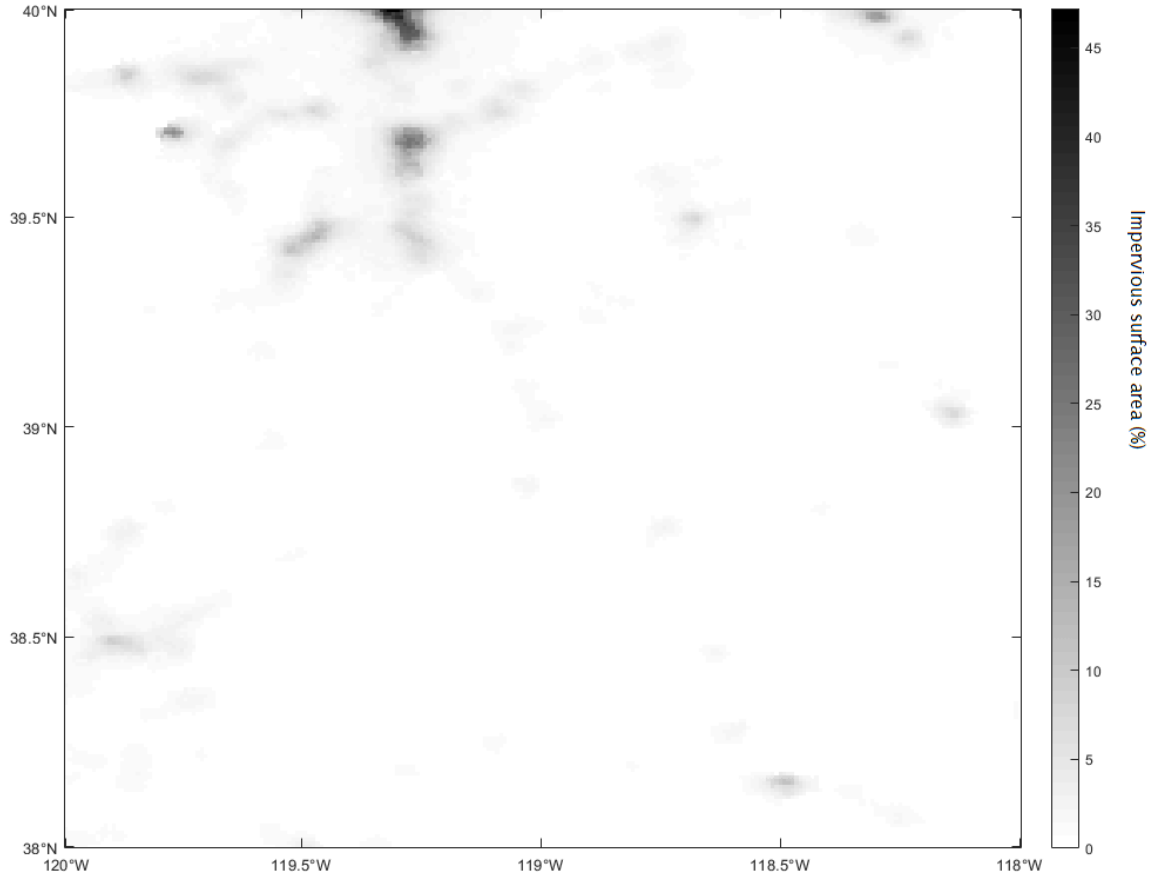


Figure 4. 5 - Representation of urban areas within the modelling domain.

4.1.1.6. Vegetation dataset

Similar to the urbanisation dataset, the vegetation dataset is important as it represents the effect of vegetation on the DEM. The original data comes from NASA's MODIS mission, with a dataset of 'vegetation continuous field' (VCF) arrays. These arrays record information about the tree density per cell. Through a corresponding global mapping of canopy heights (Simard et al., 2011), it has been possible to algorithmically derive forest heights from the VCF arrays (O'Loughlin et al., 2016). Figure 4.6 shows the anticipated height to be removed from the

DEM because of vegetation. The tallest forest heights correspond to the left side of the Sierra Nevada Mountains, and this area is the location of Stanislaus National Forest, indicating that this method is functional. Further information could be derived from the vegetation dataset about catchment response and could represent an area of future refinement in the Bristol GFM. But for now, the main use of the vegetation is for the purposes of modifying the DEM.

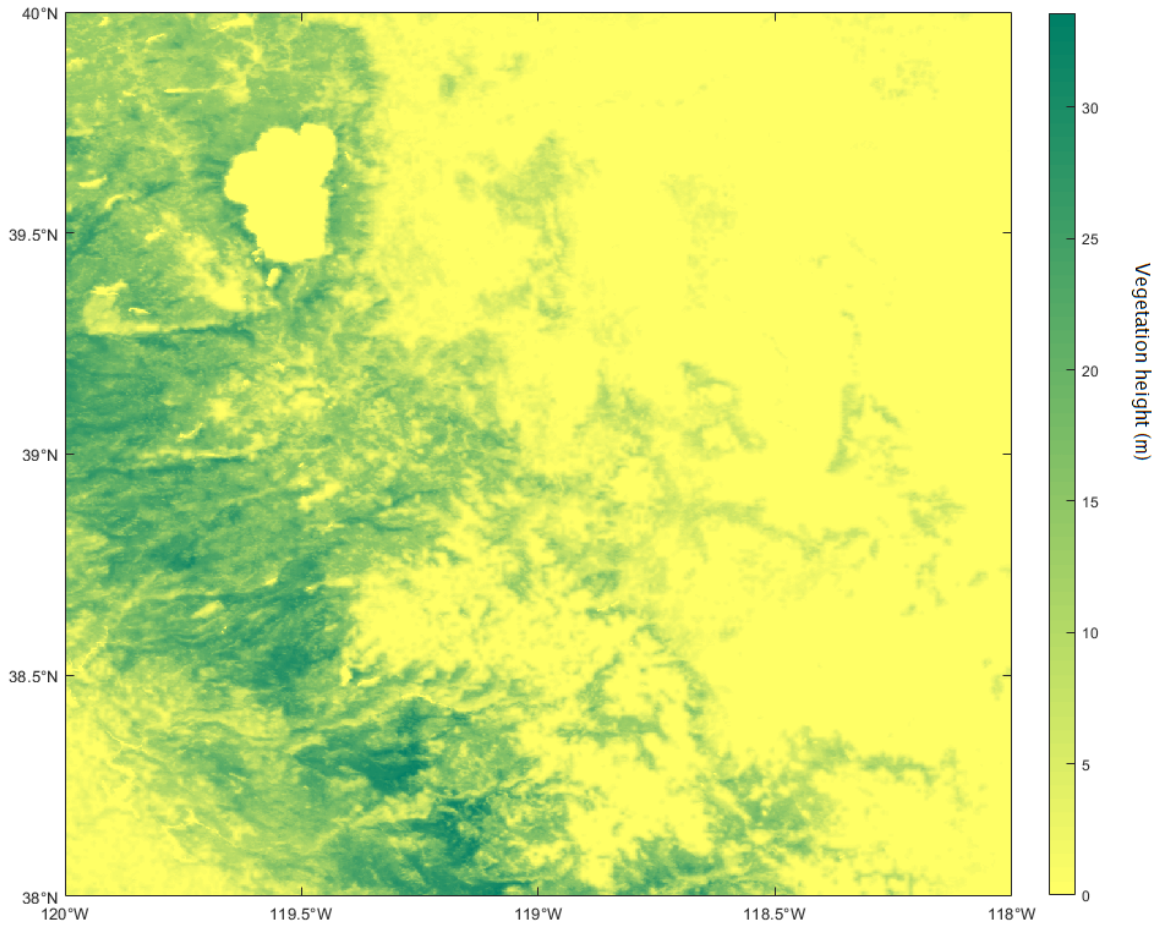


Figure 4. 6 - Representation of vegetation within the modelled domain. The darker the colour green, the denser and taller the vegetation is.

4.1.1.7. Koppen Geiger

The Koppen Geiger climate classifier has little interaction with the pre-processing elements of the model. Instead, it is used for the regionalised flood frequency analysis component for the model to generate flood peaks (Smith et al., 2015). Based on which climate type an ungauged catchment falls into, it will be estimated using only the catchments in the identical climate. While the Koppen-Geiger system has many sub-climates, the data for the Bristol GFM only classifies each 0.5° by 0.5° tile as either tropical, arid, temperate, continental or polar. In

the case of figure 4.7, there is the existence of both arid climates and tropical climates. The arid region corresponds to the arid areas of Nevada, while the west side of the Sierra Nevada mountains correspond to the climate of California, which is better classified as temperate. A clear issue with the use of the Koppen Geiger in this way is that the coarse resolution of the climate set means that there is potential for misclassification at the border between different climate zones, as is the case here, as well as local-scale variances being ignored.

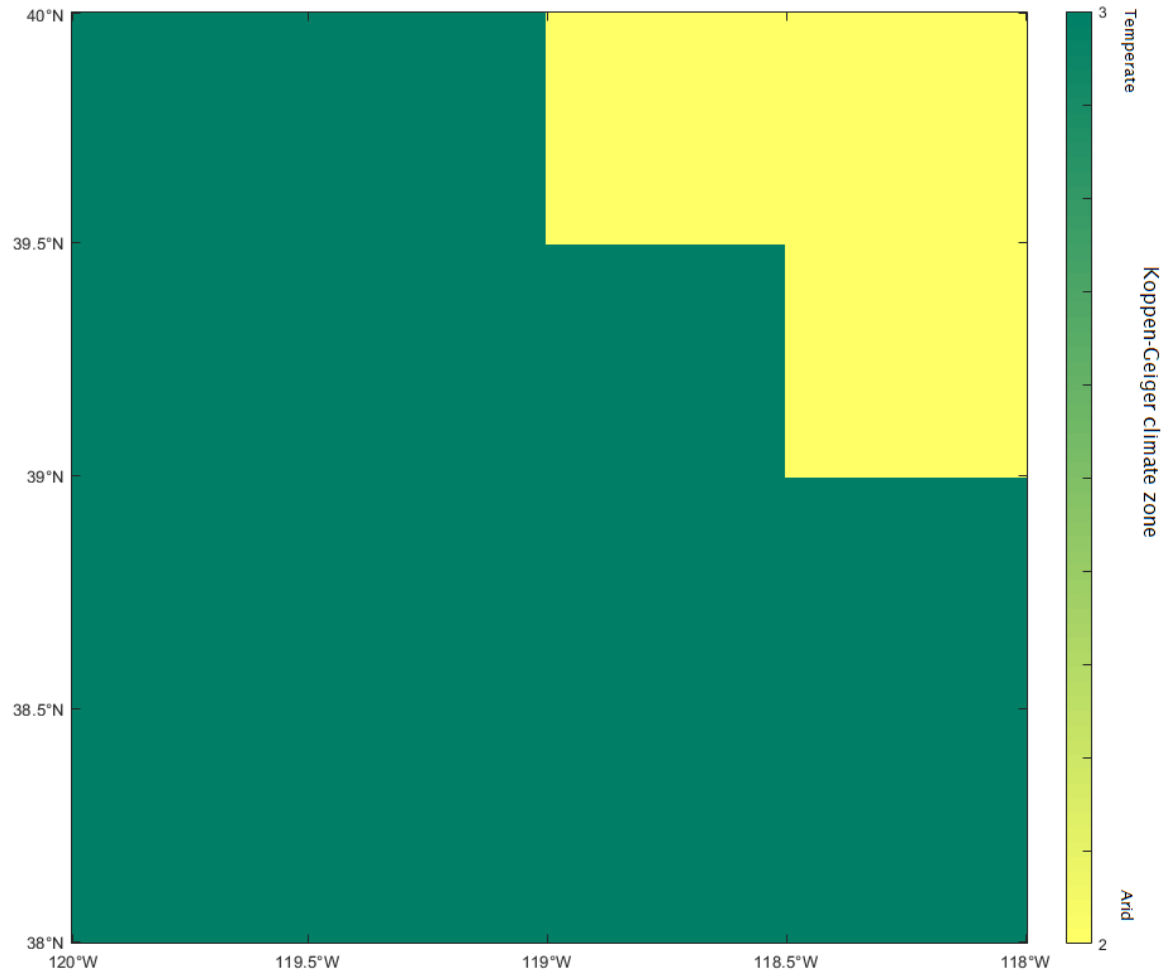


Figure 4. 7 - Koppen Geiger climate classification of domain.

4.1.2. Discharge Data

Although the primary target for the running of the Bristol GFM is its applicability to ungauged basins, it still requires some flow information from gauged basins, because of its reliance on the regionalised flood frequency analysis. The rFFA approach is dependent on an external dataset from which to pool data. This data is derived from the GRDC, which maintains a database of stations and their runoff time series. Given that the interest here is in flood peaks,

the data used is the annual maxima (AMAX) time series. While the use of catchment descriptors clusters the data, the AMAX series are used to derive probabilistic flood magnitudes for ungauged basins. Despite its utility, there is potential for criticism, given it doesn't describe anthropomorphic changes to the catchments in the dataset. Furthermore, there is great variability in the length of the time series, which means that when the procedure of "swapping space for time" takes place in the pooling, some catchments are more heavily weighted from their number of datapoints than others. The dataset is also skewed in its density as the majority of catchments are in the United States or Europe, which make the dataset a less than ideal global donor. However, given that the method is reliant upon gauged basins, it simply requires their use until developing nations have a denser gauging network.

To obtain an "observed" estimate of Q_{100} values from the AMAX time series, an extreme value distribution is needed. While the Bristol GFM pools data to overcome to issues of data length for pooling, the fact remains that Q_{100} values are difficult to estimate, because the preferred periods of record (at least 100 years) are extremely rare. That means that the observed GFM estimates we use are themselves also uncertain. In chapter 7, there are 2 regionalised flood frequency analysis methods, which we compare in terms of their predictive skill and uncertainty. However, the observations to which we are comparing them are also subject to uncertainty. A discussion of this uncertainty is covered in the next section.

4.2. Case Study Validation Data

In chapters 6 and 7 there is external data which is used with the model for the purposes of validation and refinement; which is briefly discussed here.

4.2.4. Po river data

Throughout the second results chapter, the model is tested at the Po river, requiring local data against which to examine performance. By using locally available data sources, we can see how well the Bristol GFM reproduces local events and hazard extent.

4.2.4.1. Basin authority area risk maps

To look at the model's skill in representing the hazard extent, the model was evaluated against the hazard maps of the Po river basin authority (AdBPo, 2012). While ostensibly a single 100-year return period hazard map at the local scale, the map itself is a composite of multiple

smaller scale hazard assessments, models and study areas which are the best estimates of local municipalities. This means that the hazard map itself is a combination of multiple different sources, meaning that quantifying any uncertainty in the local estimates provided is beyond the scope of this study. We are therefore slightly limited in the conclusions which can be drawn based on this data, as rather than making a comparison to actual observations, we are relying on the accuracy of the local methodologies which are employed by AdBPo. However, as the return period event can never be said to happen, the best a global model can do is replicate the results of a local model. This problem of validation for GFMs is noted in the literature (Alfieri et al., 2013; Sampson et al., 2015), where for example benchmarking to the Environment Agency flood hazard maps is the comparison of one model to another (Environment Agency, 2010). In particular, Sampson has remarked that in all cases, global scale hazard modelling must be reliant on this "high-quality benchmarking flood hazard information" in lieu of actual ground-truth, due to the nature of hazard information itself, as an ultimately statistical value. This was seen in the subsequent validation of the Bristol GFM to the FEMA flood maps of the United States, which found that the local estimates of hazard were from a variety of disparate sources, rather than being a single unified hazard estimate (Wing et al., 2017) and has proven to be a constant challenge in the assessment of models at the continental scale (Bates et al., 2020).

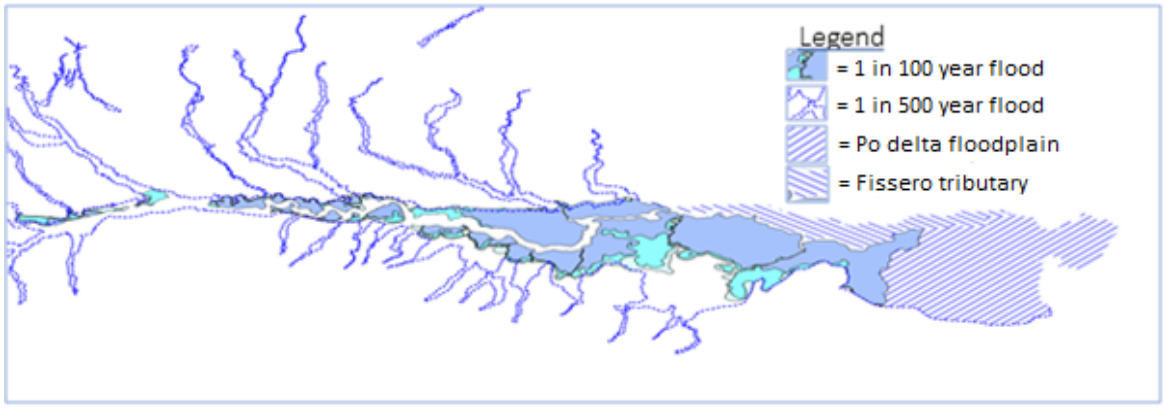


Figure 4. 8 - Local hazard "data" used by AdBPo by quantify the risks of hazard at the Po river.

Figure 4.8 shows various outlines of flood hazard which are used by AdBPo. There are areas of floodplain delineated to the delta of the Po river, as well as the Fissero tributary. These however only represent an expected extent, rather than any indication of the depth. Another hazard indicator is the blue dotted line, called the "C-buffer". This outline gives the hazard of the 1-in-500-year event, which inundates many of the incoming tributaries to the Po, as well

as the entirety of the Po river. The most important outline however, is the main 1-in-100-year floodplain, contained within the C-buffer. There are 3 layers associated with this outline, corresponding to anticipated depth, ranging from less than 1m, between 1 and 2m, and greater than 2m. While this hazard layer will be effective for comparing extent with the GFM, these depth categories are too coarse to test our model against.

The fact that the depth values of the local benchmarking data has been reduced to the precision of 1m depth tells us something about the nature of the errors in the models, even if we are not able to properly quantify them. As the shapefiles show a hard boundary between a flood depth of 1m and 0m, the implication of this is that there could be some uncertainty at the precise edge of the flood hazard layer. The only reason why this would be otherwise could be related to the levee system at the Po, which could be the constraining factor for this hazard layer. Further downstream, the uncertainties in the local model are likely increased, as the contributing factors of the Fissero tributary and the Po delta to flooding extent are also estimated, but only in terms of area. This would imply that the local model error is greater at these points, as the uncertainties are compounded by a larger number of factors, such as the inclusion of coastal flood hazard and a significant tributary.

Further upstream, we could expect these uncertainties to reduce, as the flood hazard is due to exclusively fluvial processes. As the hazard map is a collection of more localised models and estimates, we could expect that in individuals cases, these flooding extents would be quite accurate. However, given the nature of the levee system around the Po, we would not expect a singular 100-year flood event to happen across the entire basin. In such a case, the amalgamated hazard maps of the Po basin could be considered as an overestimate, as the 100-year event occurring everywhere in the basin simultaneously is arguably a rarer event overall. These considerations need to be kept in mind when the Bristol GFM is being compared to the Po hazard layer, as it has implications on how the GFM's performance can be interpreted.

4.2.4.2. Gauging station time series

While the AdBPo map can be used to measure the extent of flooding, to understand the hydrodynamics of the model will require information about the surface water elevation, which is more precise. The solution to this is to use the gauging station data along the Po river. There are 5 river gauges along the Po used in this study, each with a minimum of 18 years of

daily data. Alongside the daily flow data is the information of the river depth and elevation, meaning that a stage-discharge relationship can be derived at each of the gauge locations. By estimating the 100-year flow at each location, the corresponding 100-year depth and water surface elevation can be found, meaning that there are 5 point-source depths along the river, which can constrain the model results more effectively than the depths of the AdBPo shapefiles. Most notably, having depths and levels along the rivers as markers of model performance will tell us a lot about how successfully the model is routing discharge down the river in a realistic manner.

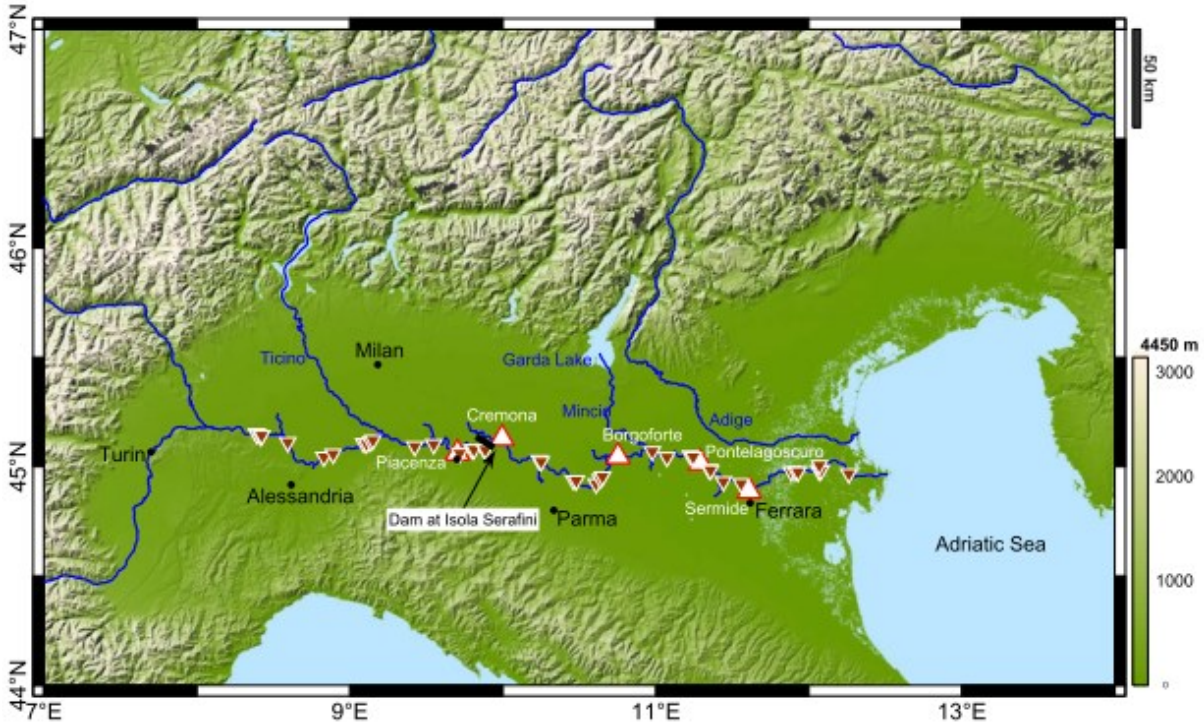


Figure 4. 9 - Local hazard "data" used by AdBPo by quantify the risks of hazard at the Po river (Credit: (Tourian et al., 2017)*).*

4.2.4.3. Profile data

The final dataset used to interrogate model performance is the bed profile depth. As explained in the chapter on the modelling framework, the depth of the river bed is a modelled variable. It is calculated through an inversion of the Manning's equation and at any river cell the Manning's n is chosen by the modeller, as well as the return period of bankfull discharge Q_{bf} and channel width. This means that the profile of the river in the model can vary substantially given the parameterisation of the model and an interesting consideration for the model is whether good performance in extent and depth comes at the cost of misrepresenting the bed profile. In this case we would be seeing a trade-off between

performance and realism, which given the number of assumptions we would expect to be quite significant.

The profile data itself is not spatially arranged. It is described as an elevation located at a point downstream, along with a description of features like an overhead bridge or nearby village. The points of measurement are also irregular. Sometimes, there are points of measurement within 200 meters, while at other points along the river there is a separation of several kilometres. Understanding how to use these data and compare it to 1km^2 resolution cells is challenging. The solution used was to first compare the data points to a Lidar DEM of the Po river, to help find the precise locations. Then points were chosen which did not have complex bathymetry, not represented in the model – this meant avoiding confluences and meanders when possible. Then the river points were cross referenced to a grid cell, so that each cell could be compared against the corresponding profile depth. From this, 23 profile points along the domain were retained and used to evaluate the model, along a river length of approximately 250km.

4.3.2. Hydroclimatic indicators

In the third and final results chapter, which looks to refine the GFM methodology, changes are made to the climate classification which the model uses. Rather than using the Koppen-Geiger method as the climatic basis for regression and clustering, a new dataset will be employed. This data set uses multiple indices across the planet to more comprehensively describe the features of the global hydroclimate relevant to extreme flow generation. These indices are the indices of aridity, seasonality and fraction of precipitation as snow (Knoben et al., 2018). There are many indices which can be used to describe aspects of the global hydroclimate. The decision to use these 3 indices over others was due to the minimal covariance between them, which indicates that each metric is representing distinct patterns of behaviour.

Each metric is based on University of East Anglia's Climatic Research Unit (CRU) climate data, drawing on distributed datasets of temperature, precipitation, evaporation, and potential evapotranspiration. These have been recorded across each month of the year, meaning that any seasonal changes in values as well as understanding how much precipitation falls as snow can be recorded. A moisture index is recorded on a monthly basis that determines whether that cell had a climate which can be characterised as water limited, or energy limited (Budyko,

1974). Summation of this moisture index determines whether the cell climate is arid or not, and the variance of the moisture index determines the seasonality. In chapter 7, there is a more comprehensive description and derivation of these 3 indices.

4.3.3. Uncertainty of Q_{100} observations

The validation of the rFFA estimates is done against estimates of Q_{100}, which is derived from the AMAX series of the GRDC. These AMAX series vary in length, meaning that when the Q_{100} value is estimated from the generalised extreme value (GEV) distributions and there is variability in the confidence intervals which are generated. In figure 4.10, we see the different confidence interval magnitudes which are produced. What is apparent is how large some of the confidence intervals are, which is a function of the very small number of datapoints in the AMAX series. Indeed, some of the confidence intervals were omitted from this graph for ease of reading. It is interesting to note that the interval sizes are comparable to a Pareto distribution.

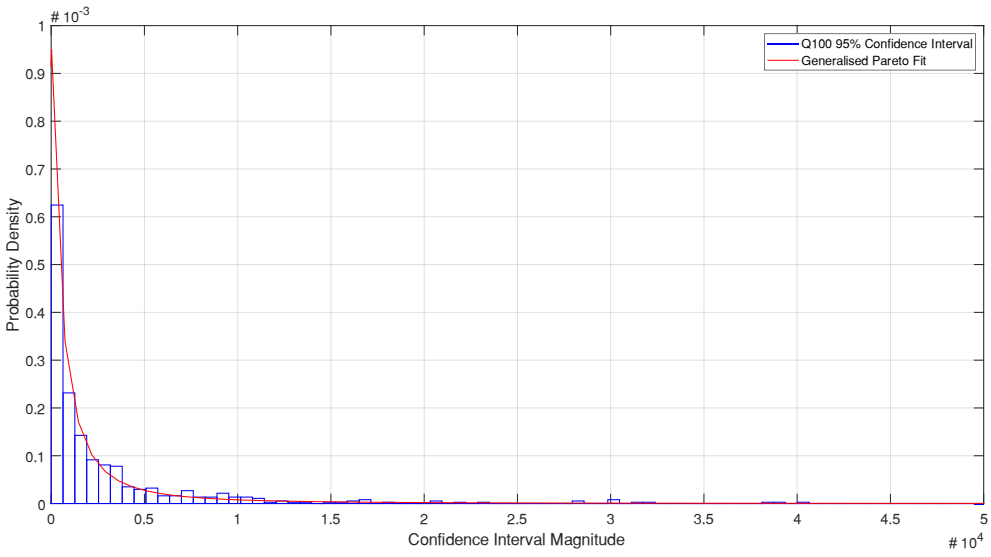


Figure 4. 10 - Histogram of the confidence interval ranges found when estimating Q100 from the AMAX of the GRDC. Data has also been fitted to a generalise Pareto distribution.

We can also look at the normalised effects of the confidence interval sizes. In the previous figure, this is looking at all of the flow estimates and the intervals are not normalised to the expected magnitude. Figure 4.11 shows the expected percentage variance across the GRDC gauges chosen for validation. This is found by normalising the confidence interval range to the estimated value of Q_{100}. Therefore, if we assume that the confidence interval followed a

normal distribution shape, this graph shows how much we would expect the value to vary by, given the quality of the AMAX series used.

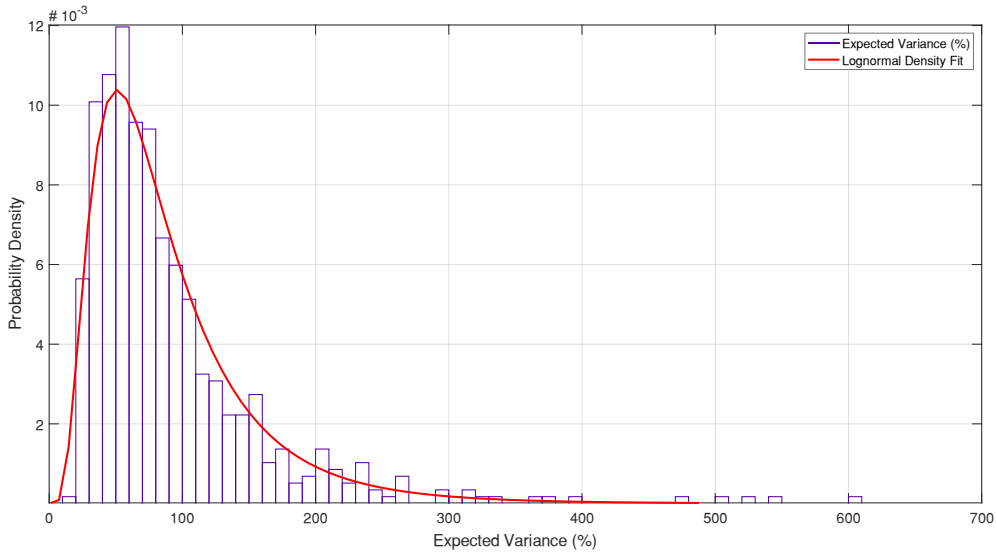


Figure 4. 11 - Expected variance normalised to the estimated value of Q100. The most common variation from the expected value is ~60%.

This small uncertainty analysis on the GRDC data underlines the difficulties faced by large scale modelling in ungauged basins. However, these datasets are currently the only tenable method to validate the global effectiveness of the GFMs. In chapter 7, we can compare the variance in the observations to the errors in the estimates from the flood frequency analyses, and see whether they are broadly comparable. If they are of the same order of magnitude then it would imply that the majority of the error in the estimates is traceable to the sparsity of data in the AMAX series.

4.3. Methods

Across the 3 results chapters, there were several methods of analysis and Monte Carlo approaches used, suited to the specific objective of that chapter. Here the selected methods will be presented and will be explained in greater context across each of the results chapters.

4.3.1. Monte Carlo simulation methods

In each of the results chapters, model evaluations were done in the Monte Carlo framework. This is a standard approach to model evaluation in the hydrological sciences and there is clear precedent for proceeding with it (Pianosi et al., 2016). While the desired outcome is different in each chapter, the experimental setups are all based around recording and correlating the

input and output distributions of the model, to identify processes and behaviours. The input distributions are usually parameters, which have a range of plausible values to be sampled from. Depending on the nuances of each case, the shape of the input distributions can vary, although this has to be done with caution, as the shape of the parameter input distribution can itself be a sensitive component in the analysis (Benke et al., 2008).

4.3.2. Sensitivity analysis

The first method employed in the analysis of the model is a global sensitivity analysis of the Bristol GFM. As previously mentioned, the main purpose of GSA is apportioning model variability to identifiable elements of the modelling chain. This variability can be structural, parametric or observational. While the various uncertainties of the remote sensed data have been noted, they were not considered for study. Also, various uncertainties in the model structure were not considered, and there is instead full focus on the parameter uncertainties. This is not to say that the uncertainty in the data and model structure aren't important, but rather than focussing only on the complexity of the parameters, it is a necessary first step to make tractable the problem of understanding the GFM.

4.3.2.1. Screening of the model

One of the main capabilities of sensitivity analyses is the ability to perform a screening. While the problem of quantifying sensitivity has become tractable by focussing only on the parameters, to reach a deeper understanding will still require a smaller pool of parameters. The purpose of the screening is to make sure that the pool of selected parameters are all sensitive, and that the parameters to be discarded are not sensitive. Choosing sensitivity analysis techniques for screening is important, as the computational requirement for screening is often a lot less than for realising the full rankings and sensitivities of the parameters in total. Hence while the use of screening methods can be parsimonious, this also means they are less likely to be robust (Gan et al., 2014). However, normally once the screening is conducted, further analysis strategies can be developed which focus on the key components of the model (Campolongo et al., 2011).

4.3.2.2. Morris method

One of the most effective methods for screening of the parameters in a parsimonious manner is the Morris method (Morris, 1991). Also known as elementary effects test (EET), the Morris method is a derivative based approach which is suited to dimensionally large modelling problems (Campolongo et al., 1999). Rather than using one-at-a-time (OAT), which has been shown as largely ineffective (Saltelli & Annoni, 2010), parsimonious sampling of the parameter space can offer valuable insights about the relative effects of the parameters. Morris achieves this through the calculation of elementary effects across the parameter space.

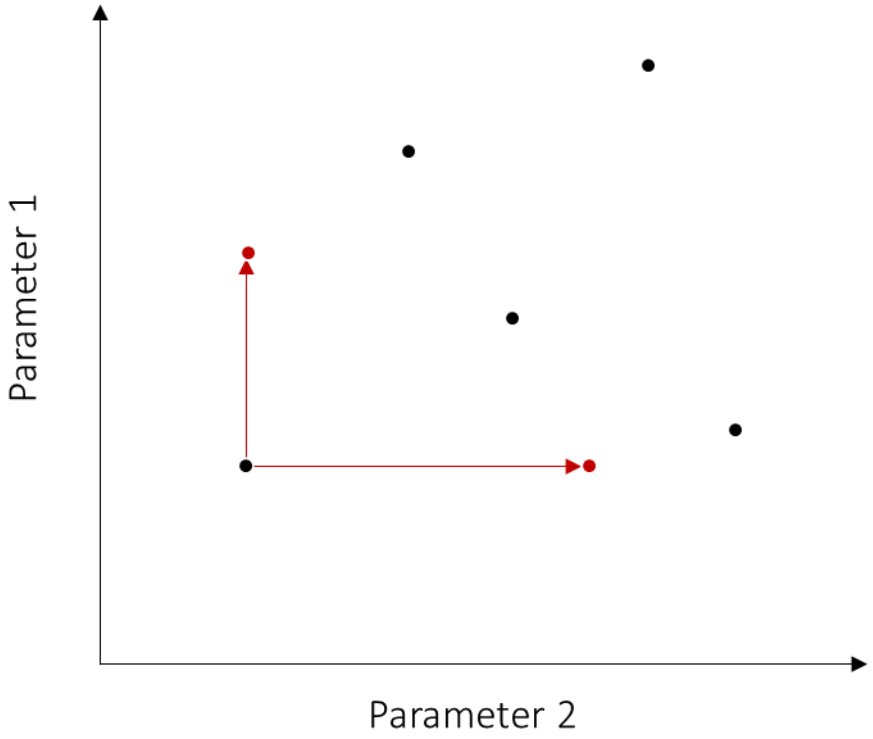


Figure 4. 12 - Simple example of Morris method. There are 5 random samples taken across the parameter space. At each sample, the ordinate of that given parameter is varied.

To help understand the concept of elementary effects, consider the two-dimensional parameter space in figure 4.12. Here, there are 5 black points randomly spread throughout the parameter space. The coordinates of each point correspond to a parameter value. Therefore, if the parameter value is varied, the corresponding coordinate of that point changes in the parameter space – represented in figure 4.10 by the 2 additional red points. Let's say it is possible to calculate a model output for each of the original 5 parameters sets. They will vary, but it is not clear if that variance is due to Parameter 1 or Parameter 2. Hence at each of these original points, the coordinates are varied one at a time, so that the difference

in model outputs is recorded per parameter. For parameter 1, the elementary effect will be recorded like so:

$$EE_{P1} = \frac{\Delta Y}{\Delta P1} \times (\max(P1) - \min(P1))$$

Where EE_{P1} is the elementary effect for Parameter 1, ΔY is the change in output due to the change in parameter space, and ΔP1 is the change in the parameter value itself. The division of these values gives the gradient, hence why EET is a derivative based method. The gradient is then normalised by the range of the parameter, so that it is directly comparable to other parameters. Notice however, that this is the elementary effect at only one sample point. Therefore, this same process is repeated at each sample in the parameter space. The mean μ and standard deviation σ of the set of elementary effects acquired is then calculated and used as explanatory indices of parameter behaviour. When μ is high, it indicates the parameter is more influential. When σ is high, it indicates that the parameter in question has some interaction with other parameters, or that its response is non-linear throughout the parameter space.

4.3.2.3. Sampling strategy

Because of the number of parameters in the GFM, the sensitivity analysis is a high-dimensional problem. This means that any sampling strategy which is computationally viable will cover the parameter space very sparsely. That means one must use a sampling strategy which can effectively represent the parameter space and extract enough information about parameter behaviour. The strategy to deliver this is the "radial design" strategy (Campolongo et al., 2011), which incorporates both all at a time (AAT) sampling strategies and one-at-a-time (OAT) sampling. The first step is to distribute the samples throughout the parameter space, ensuring they are well dispersed. Then, the OAT sampling occurs at each of the AAT sample locations. Thus, a well distributed sample of elementary effects for each parameter is collected.

4.3.2.4. Choice of outputs

As the first round of analysis doesn't focus on model skill but rather the total output variance, the choice of outputs must be signals which will best express the possible variance of the inputs. The commonplace output to use in inundation modelling is the flooding extent, which will be used here. This is the number of cells which are flooded in the simulation. However,

the success of this output is predicated on the floodplain in question being large and flat – meaning that the increase and decrease in area is clearly notable, and the signal can be easily corresponded to certain parameter values. However, for many smaller catchments this is not the case, and the floodplain can be bounded or have complex topography. In such cases the extent is less effective at describing variance. Instead, the depth of water at specific locations represents a stronger signal of model response, as the area of flooding is constrained but not the height. By using both metrics, the relative effects of the parameters will be found.

4.3.3. GLUE/Uncertainty quantification methodology

In the second results chapter, Monte Carlo will again be used to examine the behaviour of the parameters, and the effect of their uncertainty on the modelled outputs. However, there are two main differences to this method and the sensitivity analysis discussed before. The first is that the number of parameters will be reduced, given that the screening of the previous method will remove many of the parameters which are simply inconsequential (Weichel et al., 2007). The second is that in this study, the model will be assessed against other sources of data for evaluation, meaning that model skill is another component to consider. In each evaluation, the model will be assessed against the various sources of local data available, and given a score, such as the critical success index or RMSE. These scores can then be used to derive likelihood weights.

4.3.3.1. Sampling strategies

This assessment of the model will be consequently done with a smaller pool of sensitive parameters, meaning the parameter space will be sampled more densely. This is achieved with all-at-a-time (AAT) sampling, known as Latin hypercube sampling (Gan et al., 2014). The Latin hypercube strategy ensures there is good dispersion of sampling, for a given sampling density, throughout the multi-dimensional parameter space so that the range of model behaviour is fully represented. A prior of this approach is determining the nature of the input parameter distributions. What is chosen could give a different signal or model sensitivity, and should reflect the knowledge of these distributions in advance of the analysis. However, as there are no hard assumptions to account for or indeed knowledge of a-priori distributions, the prior distributions will all be uniform, which are effectively non-informative distributions. Despite this, there is still a "weighting" of the parameter values, as any value outside of a given range is therefore given a probability weight of zero (Beven, 2012), which means that

the values are therefore outside the "limits of acceptability" of the model (Beven, 2006; Blazkova & Beven, 2009). While there is no definite justification for the discontinuous weighting of parameter values which a uniform distribution gives, it is not considered problematic, because the prior distributions will be subsequently weighted with a performance metric, to return non-uniform posterior distributions.

4.3.3.2. Prior and posterior distribution weighting

The key to understanding model behaviour in this method and what constitutes a good model is the transformation of prior parameter distributions into posterior parameter distributions. As the priors for the parameters in the uncertainty analysis of the model are uniform distributions, we assume non-informative parameter distributions. Posterior parameter distributions will emerge by assessing the performance of the model distributions, such as by the use of the critical success index (Alfieri et al., 2014). Figure 4.13 shows the ideal response of the posterior parameter distribution after the likelihood weights have been added.

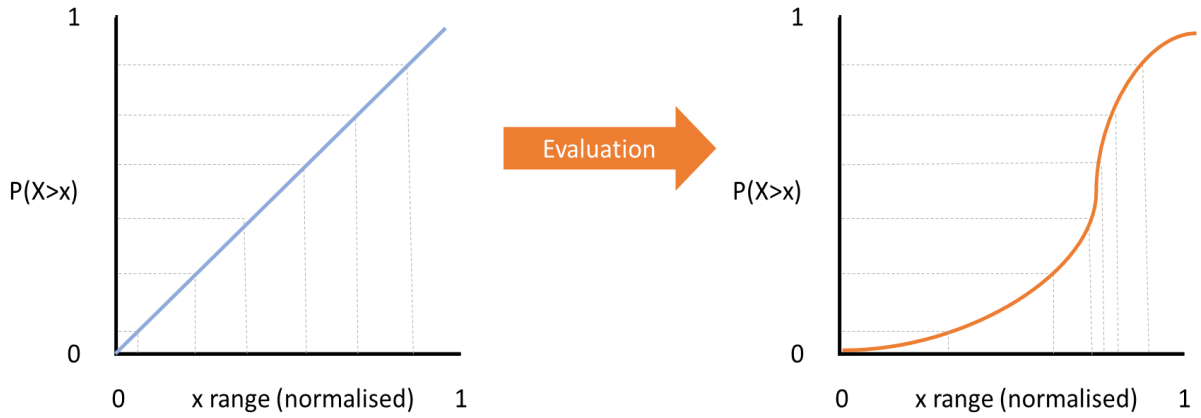


Figure 4. 13 - representation of a prior uniform parameter CDF becoming a posterior parameter CDF, due to likelihood weighting from model evaluation.

The weighting of the posterior parameter distributions requires a performance-based metric to be calculated. The precise computation of the weight will vary, as in some instances the subjectively best performance will have a minimum value, such as RMSE, while in another case the best performance will have a maximum value, such as the critical success index. The importance here is that there should be some justification that the performance metric is useful to the objectives of the modelling study and thus what the model should be representing well. Taking the critical success index as an example, the likelihood can be found in the following manner:

$$L_j = \frac{CSI_j}{\sum_{j=1}^{N} CSI_j}$$

Where CSI_j corresponds the critical success index of model j, and L_j corresponds to the likelihood of model j. Because the CSI is divided by the sum of the CSI of all the models, this means that the sum of all the likelihoods equals 1, meaning that this can be used as the posterior parameter distribution.

4.3.3.4. Derivation of prediction intervals

In chapter 7, the final prediction intervals are derived from an output distribution. In that chapter, there is a detailed derivation of the methods used to generate the distribution as well as the prediction intervals. By using Monte Carlo based methods, the end result is a distribution of model outputs, each having a conditional probability associated with them, which are potentially non-Gaussian (Freer et al., 1996). In these cases, finding the confidence intervals is still possible empirically, as the 95% range of values can be used. These will be values from the posterior distribution, weight according to the performance metric. It does however require the assumption that the output distribution produced for the data is fully representative.

4.3.4. Regionalisation of gauged flows

The process of regionalisation is based on some hard assumptions about the relationship of domain characteristics to processes and mechanisms within the domain. Within the Bristol GFM, these assumptions are held within the rFFA (Smith et al., 2015). The assumptions made in the rFFA are that data should only be pooled together when they are climatically similar, and that once this data is pooled, the only necessary explanatory variables needed are upstream area and average annual rainfall.

In chapter 7, the methods of clustering, regression and classification are used to achieve an alternative scheme of regionalisation. In the following sub-sections, brief descriptions of each method are given. For clarity, this is separate from the detailed implementations given in chapter 7.

4.3.4.1. Clustering

As mentioned, a key determinant for the regionalisation is use of the climate zones. In the original methodology of Smith et al., 2015, the climate zones were already deterministic, as the Koppen Geiger zones have set boundaries. By transitioning to a climate classification with continuous indices, there is the need to define boundaries in the climate to distinguish one zone from another. This was done using a k-means clustering algorithm. This is a method which distinguishes clusters within the pool of stations taken from the GRDC. As the new climate method has 3 indices attached to it, these can be represented as coordinates in a 3D space. Hence the stations are divisible into climate zones based on the 3 descriptive climate indices. The test for the effectiveness of the indices is measured by the degree of membership of each station to their selected cluster. This is important because it shows how many of the stations could be reasonably considered as belonging to a different cluster. However, the use of membership to determine the quality of the clustering analysis needs to be carefully considered, as the clustering itself is a mathematical construct, which doesn't really represent the processes. Use of clusters is a pragmatic choice rather than being the most realistic representation of different climate zones.

4.3.4.2. Regression

Once each of the clusters has been defined, the peaks flows can be calculated from its constituents of the MAF and the growth factor. The growth factor is calculated using a GEV distribution of AMAX peaks. However, the MAF needs to be calculated from the clustered stations. In the original method, this was done using two separate regressions, the first of which assumed a power relation between catchment area and MAF. The second assumed a linear relation between average precipitation and the log of MAF. These have some validity, but their essential goal was to ensure a constantly increasing relationship of MAF with both area and precipitation. However, this can also be achieved by fitting a multivariate regression, with both precipitation and area as predictor variables. Using a least-squares fit optimiser, MAF surfaces were developed for each individual climate cluster. By contrast to the MAF surfaces developed for the other Koppen Geiger clusters, these surfaces showed characteristic variation in the shape and slope of the surfaces, indicating that they are representing climate with genuinely different processes.

In all cases, the curves and surfaces were fitted using a least-squares optimiser. In this, the parameters of the regression fit were varied until the sum of squared errors was minimised. Alongside the best fit, information about the goodness-of-fit was also computed, which could then be used for the analytical derivation of the prediction intervals as mentioned above and fully described in chapter 7.

4.3.4.3. Classification

While clustering and regressions are performed on the training data, it is also necessary to train classifiers on the test data. The clustering of the climate zones occurs without knowledge of the test data, so they are not assigned a climate cluster at the beginning of the analysis. Therefore, it was necessary to train classifiers which could then assign each test station to an appropriate climate cluster, as well as the correct growth curve sub-cluster. These classifications were done using a fine-tree classification method. Fine-tree classification is a machine learning technique which learns about the boundaries of the data by producing bifurcating decision branches, which ask whether this value is higher or lower than the threshold, for example. The fine tree algorithms were all trained successfully, all of which exceeded an accuracy of 99%, with the only errors occurring at the borders between clusters.

5. Research Chapter 1- Sensitivity: Sensitivity Analysis of a Global Flood Model

5.1. Introduction

This results chapter is formed from components of a pending article on the basic behaviours of a GFM and marks the beginning of analysis. It is necessary to have a section which examines the relative sensitivities of all GFM parameters, as in the current form, with no prior knowledge, the sensitivity of the model to the uncertainty of the parameters is unknown. Parameters have been chosen as a key area for examination, prioritised instead of the effects of data uncertainty and structural error. One of the reasons for this choice is that many of the parameters exist to account for the uncertainties and inaccuracies in the global datasets. This means that by extension, where there are parameters concerning the global data, it can be expected that some of the uncertainty of the input data has been accounted for.

As discussed in the introduction, the development of GFMs is continuing at a steady pace, although the discipline has not yet reached the stage of formal inter-comparisons of methods. Confirming this, there are studies which highlight a lack of shared knowledge between each GFM research group (Bernhofen et al., 2018; Hoch & Trigg, 2019), which point to the main challenge of this emerging discipline being a lack of understanding of the relative strengths and weaknesses of the various modelling approaches.

Alongside this lack of established methodologies, there has been little investigation into the detail of how these models work, and what differentiates them from their local-scale variants. While the differences of datasets used, domain scale, and process representation are clear enough, the application of GFM's to ungauged basins further complicates matters, as the effectiveness of each model is only understood in terms of its comparison to local data, which is contextual. However, to gauge the relative effectiveness of the models, it is first necessary to understand the behaviour and effectiveness of the models independently, as it will then follow that their structural differences can be compared.

It is in this context that the following chapter presents a contribution to the new field of global flood modelling with the examination of sensitivities and associated uncertainties of the Bristol GFM, with no prior expectation of what they should be. The analysis is to be wholly exploratory, meaning that the "performance" of the model is not assessed, only their raw variability. While basic in approach, this investigative work is a necessary step for the

continued development of GFM's, and the results found herein will form the basis of the following chapters.

This study is concerned with the GFM developed at the University of Bristol (Sampson et al., 2015), which will be subjected to a global sensitivity analysis (GSA). GSA is the "study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input" (page 1, Saltelli et al., 2007). In a general sense, the inputs of a model can have uncertainty in the parameters, data, and the model structure. In the case of the Bristol GFM, this produces an high dimensionality problem, and one which appears intractable, if the uncertainty of the parameters, data, and the model structure are all considered simultaneously. Hence the problem needs to be simplified somewhat. Rather than focus on every classical form of uncertainty (Götzinger & Bárdossy, 2008), this study will limit itself to only the parameter uncertainty, itself remaining a dimensionally large problem to tackle. This means the following sensitivity analysis will consider all of the sources of parameter uncertainty in the model, which are implicitly linked to the remote sensed global datasets on which the model is based, the underlying hydrodynamic model, and the extreme value generation, via the various pre-processor elements which make up the overall modelling framework.

5.1.1. Addressing the credibility challenge

The scientific community has been able to simulate flood hazard over small spatial domains, with both accuracy and parsimony for at least 20 years (Bates & De Roo, 2000). It follows that global flood risk models (GFMs) could possess the capacity to estimate flood hazard, although the move from local to continental scale data means there is a corresponding change in model structure and parameterisation, which requires identification. There is a risk that the shift to continental scale inundation modelling could mean a divergence of modelling outputs, as the number of assumptions and possible approaches increases. This was confirmed in a study assessing the agreement of modelling results across the entire African continent (Trigg et al., 2016). In favourable locations such as well bounded floodplains around large perennial rivers, it was found that the models were largely in agreement. However, across environments with more varying processes, particularly arid and deltaic regions, the GFMs diverge in their estimation of hazard, with model agreement reducing to 30% by flooded extent. This leads to

what Trigg et al. termed the "credibility challenge" global flood models. While it is very unlikely that all of the models will ever be in perfect agreement, without understanding the underlying reasons for the variability in outputs, the credibility of the entire discipline will remain in question.

What is driving these disagreements between the GFMs? We know principally there are structural differences between the models (Hoch & Trigg, 2019); the various research teams have applied different methodologies, such as the use of a 2D hydrodynamic model, or a 1D routing model. There are also differences in algorithmic representations of river networks, for example some models produced a comprehensive river network across the Sahara desert, which while proven to be real, are no longer the recipients of rainfall and flow (Coulthard et al., 2013). If the modelling structures were totally analytical, these structural differences would be the only source of disagreement. However, due to the dependence on remotely-sensed inputs and simplification of process representation in these models, the resultant parameterisations present additional sources of uncertainty in the GFMs (Yamazaki et al., 2014). These uncertainties are an unavoidable component of all flood modelling but are magnified when attempting to model at the global scale, often in data-sparse regions, often with the objective of modelling extreme events. Therefore, addressing the "credibility challenge" (Trigg et al., 2016) of the models and improving simulation accuracy depends upon addressing these accompanying uncertainties of the modelling structures, parameterisations and inputs, and developing our understanding of how they are propagated through to the modelled hazard layers (Hall *et al.*, 2005; Ward *et al.*, 2015; Savage *et al.*, 2016).

5.1.2. Understanding model structure

Sensitivity analyses have been performed on flood models previously (Hall et al., 2005; Pappenberger et al., 2008), however only partial sensitivity analyses of GFMs have been performed to date (Dottori et al., 2016). The application of CaMa-Flood GFM (Yamazaki et al., 2011) to river deltas (Eilander et al., 2018b) is one example, which involved assessing the effect of including river bifurcations in floodplains and coastal regions. This study addresses the impact of the structural difference due to various modelling choices, such as river bifurcations, have on flood extent. Meanwhile, the sensitivity analysis of Dottori et al., 2016 addresses model parameter sensitivities in a one-at-a-time format, therefore only providing simple explanations of various uncertain components. These studies represent a sensitivity

analysis of the functionality of singular components of the model varied one-at-a-time. While this is important, it does not allow us to look at the "bigger picture" of model behaviour. Our objective is to understand which sections of the parameterisation scheme are most influential on the model output and to what extent they interact with other parameters. To this end, we must consider all of the parameters and vary them simultaneously.

We are also interested in how the sensitivity of the model output to parameters might change with location. It seems a clear intuition that the flooding of a flat arid domain will not behave in the same way as a tropical hilly region, and there is variability in several processes, such as domain topography (Manfreda *et al.*, 2011; Samela *et al.*, 2017) and extreme rainfall (Berghuijs *et al.*, 2016) which can produce an array of mechanisms, which can affect an extreme event at a given location. We are interested to see whether the corresponding peak flow generating mechanism discussed in section 3.3 is able to capture this variability. If it is capable, then we would expect a different response from the parameters associated with the extreme value generation of the GFM, when applied to different geoclimatic locations. This is an established discussion in hydrology (Wagener *et al.,* 2007) and the subsequent IAHS PUB decade of research testifies to its central importance.

In the Bristol GFM, the definition and behaviour of the parameters is often a function of the datasets which are being used. As the global model is reliant on globally available and consistent datasets, then any global methodology and parameterisation needs the flexibility to accommodate the variability of the entire global hydrological and earth systems. Furthermore, when one considers the uncertainties associated with the remotely sensed datasets, such as SRTM DEM (Rodríguez *et al*, 2006), or the discrete representation of climate in the Koppen-Geiger classification used to inform extreme flow generation (Peel *et al*, 2004), it becomes apparent why there are so many parameters (see section 3.6). What is not yet clear, is whether the number of parameters used is optimal, and if the current parameterisations have the correct level of flexibility to represent the global hydroclimate. Similar tension for more local scale hydrological modelling has been noted by Van Werkhoven *et al.*, 2008:

"A fundamental trade-off exists in watershed modelling between a model's flexibility for representing different watersheds with different characteristics versus its potential for overparameterization."

Watershed representation is clearly different across each of the contemporary GFM frameworks (Bernhofen et al., 2018), in that the cascade of modelling decisions and how the watershed domains are formed in each model is different. But in a practical sense, the processes are all similar. There needs to be the definition of a domain over at least the continental scale; there needs to be a usable hydrography within this domain to represent river size and location, and there has to be a forcing of extreme flows. This adds up to the "from data to hazard" perspective mentioned in previous chapters. While all of these do vary across the GFM's, many of the problems probably exist in the corresponding module of other GFM's, meaning if there is a particular aspect of the modelling chain which is highly uncertain, the same aspect of the other GFMs should also be examined.

It is worth highlighting here that although the analysis of uncertainty is being limited to a sensitivity analysis of the parameters, due to the structure of the Bristol GFM, many of these parameters are completely novel, and emerge through the necessity of reformatting the various large-scale datasets to the task of inundation modelling. Hence when there is large sensitivity about one of the parameter values, this can actually be informative about the uncertainty of the related data which the parameter interacts with. In this sense, the following sensitivity analysis is mostly beneficial as a heuristic tool for understanding general tendencies of the model, while obtaining specific information about the parameters.

5.3. Methodology

A screening of the parameters via sensitivity analysis (Morris, 1991; Campolongo *et al.*, 2011) will be performed in this study, with the aim of identifying the division between parameters which are sensitive and insensitive. Doing this will address whether the Bristol GFM fulfils the criteria for being overparameterized. It is important to test this over several regions, as the behaviour of the model at a single watershed might not be indicative of general parameter behaviour (Beven, 2000).

It is necessary to begin testing of the GFM in gauged basins before any confidence of modelling ungauged basins can be had. This is because any confidence we can have about model performance in ungauged basins has to be extrapolated from any confidence we have within gauged basins, and the transferability of this skill. Data-rich regions such as Europe and

North America have climate zones understood to be mostly temperate and semi-arid, with some continental areas. While this is a limited spread of climate zones, it ensures that the modelled domains can be used in future uncertainty analysis and model evaluation studies, although in this study only the variance of the raw output is examined. The model outputs we are observing are the resultant flood inundation extent across the domain, and the depth of the flood waters. To acknowledge uniqueness of place in model domains, we conducted the sensitivity analyses at the three lower river basins of the Sacramento, Po and Mississippi, which are justified in more detail in the next section. We will use a sensitivity analysis within a Monte Carlo framework (Weichel *et al.*, 2007) to test how simulation outputs change with respect to varying input parameter sets, as a representation of parameter uncertainty. From this, we hypothesise it will be possible to screen out the majority of the 36 parameters first presented in the model structure of chapter 3.

5.3.1. Case study – selection of domains to model

To run the GFM a domain needs to be specified. Because we are limited by the time associated with the running of the models and the nature of the model diagnostics being performed, an important consideration for domain selection is the computational requirements. Being limited by the computational cost of the model, we are unable to produce continental scale domains in a Monte Carlo setup. Instead, a domain size of 1 degree by 1 degree was selected, which comes with an in-built buffer of half a degree, meaning that the domains produced have a size of 2 degrees by 2 degrees, or about 57,600 km^2 at the equator – this area is reduced as the domain moves closer to the poles. This size of domain was chosen because it ensures that Monte Carlo simulations can be performed in a period of days instead of weeks at the continental scale.

As mentioned, we perform sensitivity analyses at 3 locations. Given that each location should have local data so that it might be carried forward in future study, this limits the areas for consideration to mainly North America and Europe, and therefore consideration of the main basins therein. Furthermore, we can select sites according to how the model recognises domains as different. For example, the GFM's use of the Koppen-Geiger classification system when estimating extreme flows which limits the choice of climates within Europe and North America to temperate and semi-arid climates. As such, we have chosen the sites of the Po lower river basin, the Sacramento, and the lower Mississippi. These sites also have variability

in their topography, with the upper Po catchments coming from an alpine environment, the Sacramento being fed from mountainous regions but with the semi-arid Californian climate and the Mississippi being a very large floodplain. Beyond these topographic and climatic differences, as well as being in gauged locations, these sites are sufficiently large basins that at the continental scale, we can expect to see larger variation and events, while also having the option to assess smaller scale variability if needed.

Po Domain

Figure 5.1 describes the Po domain in terms of each of the remotely sensed data inputs, as shown in chapter 4. We can see from the figure that the Po basin is a large valley, which has the southern Alpine and Apennine mountains on either side with a total upstream accumulation of approximately 12,000km^2. We also see that most of the rainfall occurs in the mountainous regions. The only substantial water surface in the region is Lake Garda, to the north of the main river, while, the accumulation raster shows the upstream area of the Po river, as the most significant river in the region. There is a 400mm variation in average rainfall across the domain, ranging from 700mm in the floodplain to over 1100mm in the Apennine and Alpine regions. There is little vegetation in the floodplain area, as this is mostly taken for agriculture or, as we can see from the urbanisation raster, there are many significant population centres in the Po valley, given it is a historic population centre.

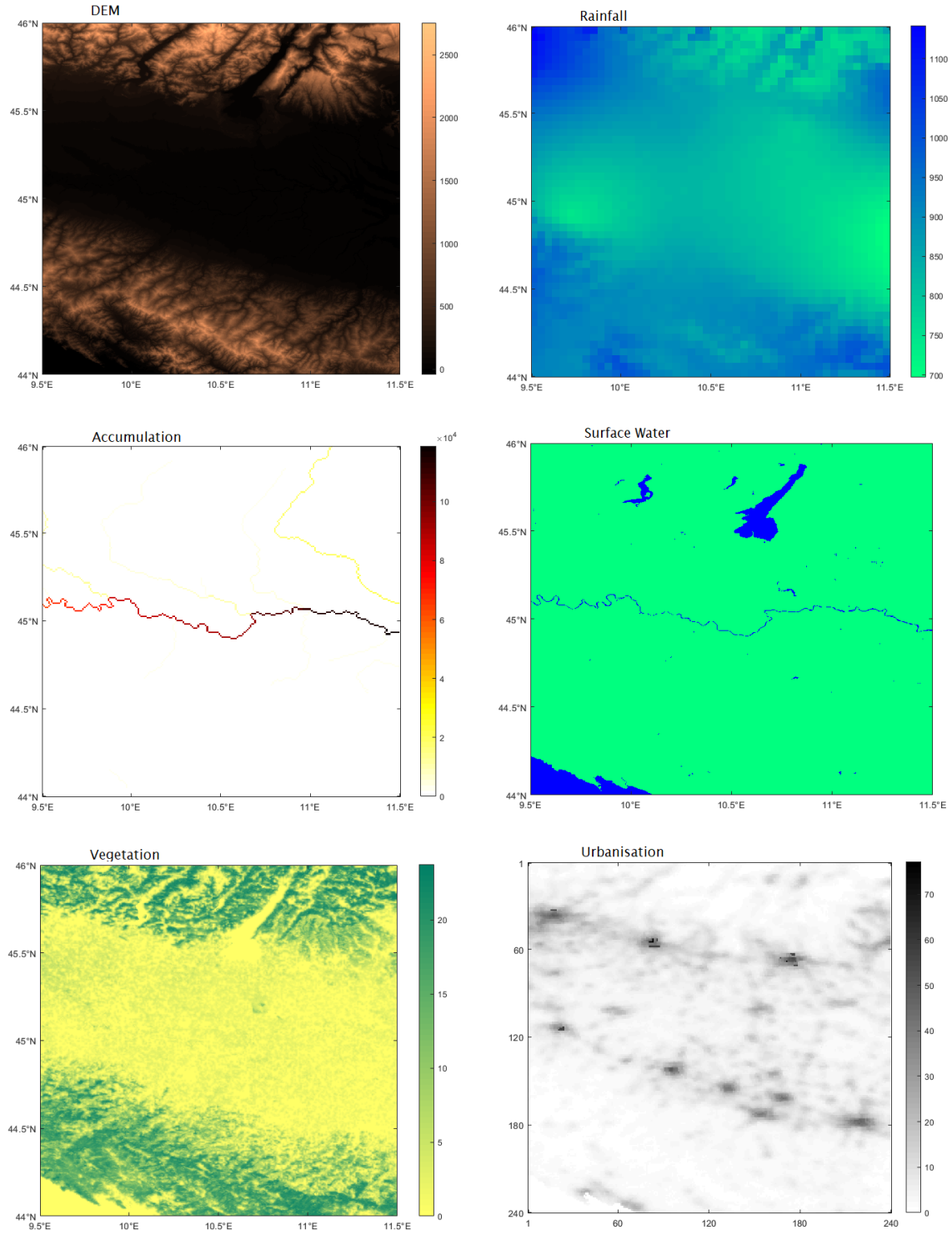


Figure 5. 1 - Remotely sensed input datasets used to describe the Po domain. The DEM panel represents the elevation in meters. Rainfall is represented in average annual rainfall in millimetres. Accumulation is represented in square kilometres. Surface water is a binary class. Vegetation and urbanisation are represented using a percentage.

Mississippi Domain

In the Mississippi domain, the variability of the DEM is very small. There is only about 150m change in elevation over the entire domain. However, despite this there is variability over the floodplains with lots of morphological features associated with tributaries and the formation of many oxbow lakes along the main Mississippi river. This is corroborated by the surface water dataset which shows the floodplains as having many small scattered lakes, and a swampy area to the east of the main river. The accumulation metric again shows the size of the Mississippi river accumulation as approximately 4.5 million kilometres squared, to be expected given the size of the greater basin. Other points of interest are the prominent areas of vegetation along the banks of the river, as well as a city located at the centre of the domain. Rainfall is not spatially variant, averaging approximately 1400mm over the entire domain, lowering slightly to 1260mm in the NW part of the domain.

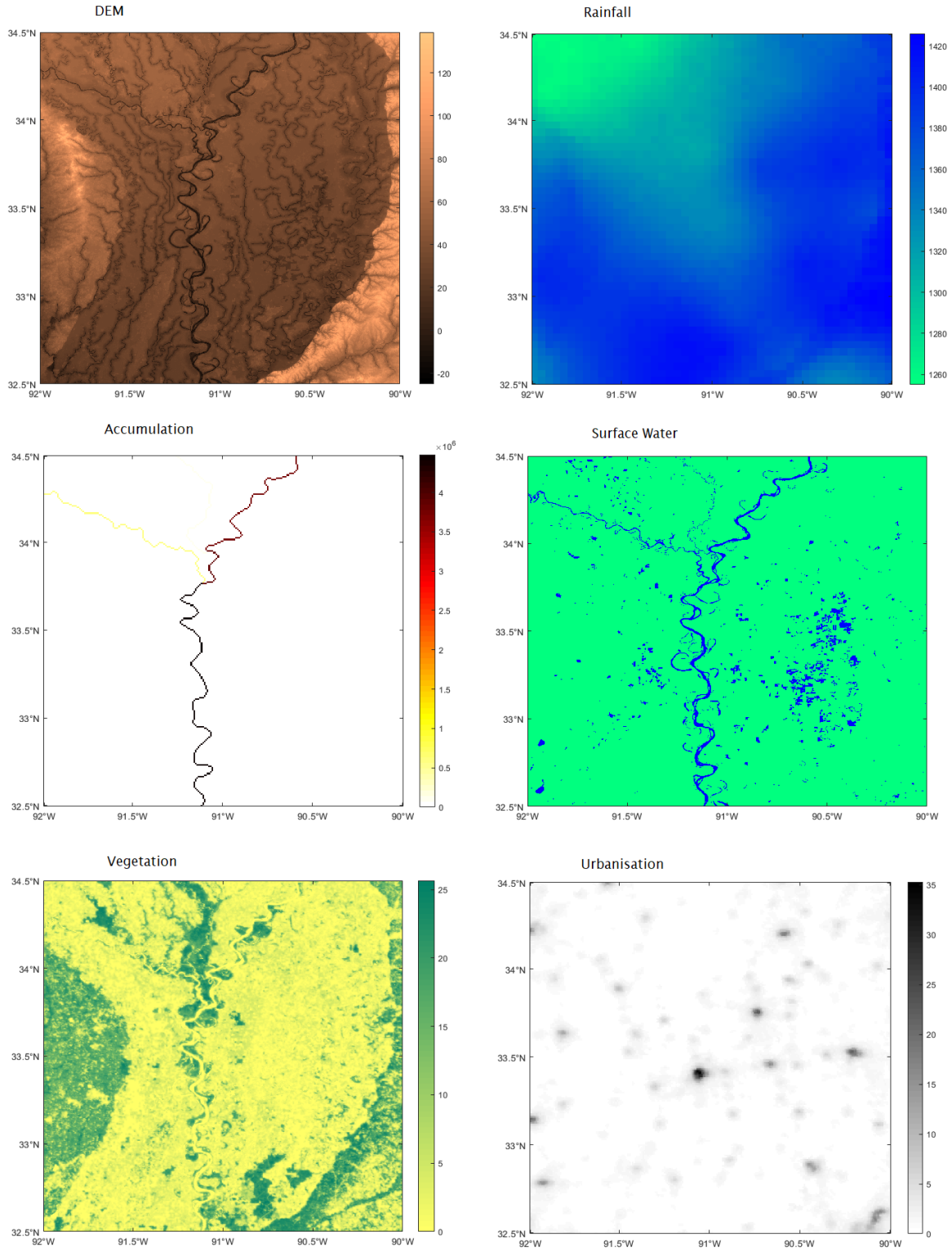


Figure 5. 2 - Remotely sensed input datasets used to describe the Mississippi domain. The DEM panel represents the elevation in meters. Rainfall is represented in average annual rainfall in millimetres. Accumulation is represented in square kilometres. Surface water is a binary class. Vegetation and urbanisation are represented using a percentage.

Sacramento Domain

The main features of the DEM shown by the Sacramento Domain is the Sierra Nevada Mountains, which trail down to the valley floor. One can see from the rainfall raster that the majority of the rain comes from the mountains with annual averages over 1600mm, although where there is some more elevated terrain on the west side of the domain, there is also average rainfall of the magnitude 1000-1200mm. The accumulation raster of shows the Sacramento river where it flows out into the Bay Area, as shown by the corresponding surface water raster, which reaches a maximum of 18,000km^2. While the majority of the vegetation in on the elevated terrain, there are several areas of very dense populations, corresponding to the San Francisco metropolitan area in the southwest of the domain, while the central area of urbanisation is Sacramento.

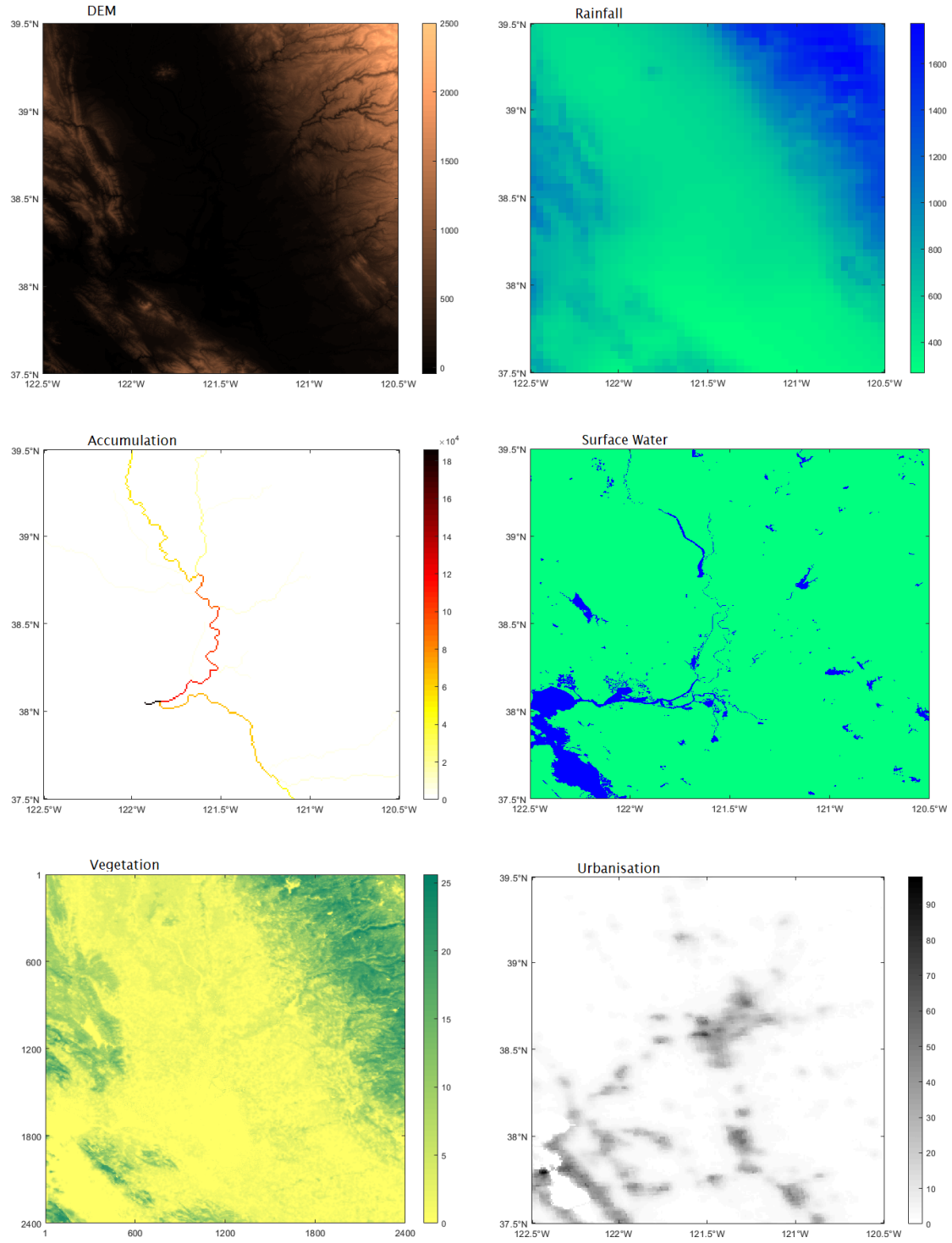


Figure 5. 3 - Remotely sensed input datasets used to describe the Sacramento domain. The DEM panel represents the elevation in meters. Rainfall is represented in average annual rainfall in millimetres. Accumulation is represented in square kilometres. Surface water is a binary class. Vegetation and urbanisation are represented using a percentage.

5.4.2. Model setup and information

In this study, each parameter identified and described in chapter 3, will be implemented as an uncertain component, so that the sensitivity of this uncertainty can be better understood.

The Bristol GFM will be set-up in a Monte Carlo methodology so that the output parameter distributions can be compared to the input distributions, with the aim of screening out parameters which aren't important and giving some first impressions of model behaviour. The behaviour to be recorded here is the raw variance of the output distributions according to the input distributions. Raw variance means that the performance of the model at the 3 locations won't be assessed, as the outputs are not being compared against validation data. By doing this, our conception of the model will not be conditioned by any information about its skill; we are simply interested in the response of the model itself.

5.3.3. Morris method and screening approach

For models with a large number of parameters, the ideal approach is the method of Morris, or "elementary effects test" (Morris, 1991), as a relatively low number of model realisations are required per parameter set. This is a "derivative-based" method, which tracks the response to individual variations in the parameter values throughout the parameter space. The method is derivative-based because the computation of an elementary effect assumes a linear effect; hence the elementary effect can be considered the gradient between the change in parameter value and the model response at a certain two points in the parameter space. By computing the summary statistics about these individual variations throughout the parameter space, indices of the importance of a given parameter can be derived. These indices inform the user about whether the parameter's effects are negligible, linear and additive, or non-linear and interactive (Campolongo *et al.*, 2007). For parameter *i* which changes by the magnitude Δx_i at point *k* in the hyperspace, the elementary effect is determined to be:

$$EE_i^k = \frac{Y^k(\boldsymbol{X^k} + \Delta x_i) - Y^k(\boldsymbol{X^k})}{\Delta x_i}$$

Where *EE* is the elementary effect of the parameter, *Y* is the output of the model, $\boldsymbol{X^k}$ is the model input of parameter set $\boldsymbol{k}$, and Δx_i is the perturbation of parameter *i*. Hence for each parameter *i* there is a distribution of elementary effects, F_i. The summary statistics used in this approach are the mean and the standard deviation; termed μ and σ respectively. For a distribution containing *r* elementary effects, these statistics are computed as:

$$\mu_i = \frac{\sum_{k=1}^{r} EE_i}{r}$$

$$\sigma_i = \frac{\sum_{k=1}^{r} (EE_i - \mu_i)^2}{r}$$

A higher mean implies a greater overall influence of the parameter on the output, while a greater standard deviation indicates that the effects of the parameter could be non-linear, or rather interact with other parameters (p. 111, Saltelli et al., 2007).

Our sensitivity analysis is being performed by using the SAFE Toolbox (Pianosi *et al.*, 2015) in MATLAB, with the aim of screening the parameters to discover which are unimportant. A secondary objective of the study aims to see if there are a subset of parameters which are recurrently important, which will become the focus of the later chapters in the thesis.

5.3.4. Selection of parameter ranges and sampling

The Monte Carlo approach requires each parameter to have a range from which to sample, as a representative of uncertainty. As the GFM's are globally applicable, and the results of the sensitivity analysis are to be globally applicable, then it follows that the ranges should be equivalent to the global range of these values, meaning the greatest feasible range for anywhere on earth. This is particularly true for simulation of the GFMs, as in the application to ungauged basins it is the case that there are no constraining observations. Definition of the ranges was difficult, given the non-physical basis of many of the parameters. In such cases, ranges required estimation. This meant allowing deviations from the default parameter values, which wouldn't make the parameter falsely sensitive. As such, the deviations were no larger than the order of magnitude of the parameter itself, and normally the range was ±50% of the default value. This is a nominal deviation, which attempted to provide a suitable level of variance in the parameters for it to be noticed in the sensitivity analysis, while not exceeding the limits of acceptable behaviour, which is extremely difficult to predict in advance. In the cases where the parameters have some clear applications prior to the development of the Bristol GFM, which is principally the roughnesses as determined by Manning's n and the bankfull discharge Q_{bf}, the ranges were taken from literature, with ranges which reflect the global applicability of the model. (Arcement & Schneider, 1989; Castro & Jackson, 2001; Dalrymple, 1960).

We are uncertain what the initial parameter sensitivities will be, despite having hypotheses about the dominance of Q_{bf}, and Manning's n, due to their influence on critical model parameterisations (Aronica et al., 2002; Pappenberger et al., 2008). Therefore, the global ranges of the parameters are employed with the minimal level of constraint. This lack of constraint is due to the fact that this chapter is purely exploratory and we are not yet placing emphasis on model performance. Therefore, such as in the case of our hypothesised important parameters, we have taken global ranges to understand the totality of plausible effects, rather than truncating the ranges to something locally suitable, which would be best practice in the case of a local study. Although one risks masking the sensitivity of some parameters due to the large effect of others, which wouldn't be noticed until the highly sensitive parameters in question have been constrained, this couldn't be done with prior knowledge. This is especially significant, given there are no constraining observations and highlights the reason for presenting nominal ranges discussed previously. Furthermore, in the hierarchy of importance, if a parameter does not register as sensitive unless other parameters are set as deterministic, then this is not a parameter of critical importance.

As discussed in section 4.3.3.1, we decided to proceed with uniform sampling ranges for the parameters. This was considered to be less of an unjustifiable assumption that sampling for normally distributed ranges, given that we have no evidence for normality in the case of any parameters, meaning that the only irreducible assumption we are left with is that of uniform parameter ranges. While this means that there will be a greater overall tendency towards parameter values at the extremity, the fact that in any one circumstance there are 36 parameters means that the likelihood of implausible parameter combinations remains extremely low, and the occurrence of a small number of these outliers will not impact the results of the study. Furthermore, the radial sampling strategy (Campolongo et al., 2011; Saltelli & Annoni, 2010) was employed, meaning that parameter values were deviated independently of each other, which again diminishes the probability of extreme parameter combinations.

Parameter	Ranges	Parameter description	Position in structure
Urban def par a	0.015 – 0.095	Change size of channel at bankfull discharge due to urbanisation	Urban
Urban def par b	0.5 – 0.85		Urban
Q_{bf} (years)	0.5 - 30	Number of years for bankfull discharge to occur	Hydraulic
Accumulation threshold (km^2)	40 – 60	Minimum accumulation to define a river in model	Hydraulic/ topographic
slope threshold (m/m)	0.015 – 0.045	Slope threshold to remove DEM artefacts	Topography
Surface water threshold (km^2)	70000 – 130000	Determines if a water body is a river or other	Hydraulic
neighbour filter (km^2)	20000 – 50000	Used to derive bank heights	Topography
Bank smooth par a	0.03 – 0.3	Smoothens bank heights along river profile – maintains monotonic rivers	Topography
bank smooth par b	0.1 – 0.6		Topography
bank smooth par c	2 – 10		Topography
arid screen par - rain (mm)	150 – 450	Removal of channels in the case of arid environments	Climate
arid screen par – acc (km^2)	500 – 1500		Climate
hyper-arid screen – rain (mm)	20 – 80	Removal of channels in the case of hyper-arid environments	Climate

hyper-arid screen – acc (km^2)	10000 – 100000		Climate
manning's n	0.025 – 0.1	Determines channel roughness	Hydraulic
min reach stream length (km)	5 – 15	Minimum stream size defined for model	Hydraulic/ topographic
max reach stream length (km)	15 – 25	Minimum river size defined for model	Hydraulic/ topographic
river/stream threshold (km)	500 – 1500	Determines accumulation for stream or river	Hydraulic/ topographic
reach length max (km)	150 – 250	Maximum reach size within the model	Hydraulic/ topographic
reach acc threshold	0.5 – 0.95	Threshold determining accumulation increment per each river length	Hydraulic/ topographic
area/rainfall weight	0.1 – 1	Weighting of rainfall vs. area in regionalisation	Flood frequency analysis
min stream velocity (ms^{-1})	0.2 – 1.5	Stream velocity minimum limit	Hydraulic
meander coefficient	0.7 – 1.3	Coefficient of channel meander at the subgrid-scale	Topography
floodplain n	0.025 – 0.15	Determines floodplain roughness	Hydraulic
veg filter par a	0.095 – 0.135	Determines reduction of DEM height in presence of vegetation	Vegetation
veg filter par b	0.36 – 0.48		Vegetation
urban extent low (%)	0 - 5	Degree of urbanisation whereby cells are removed, assumed to be buildings	Urban

urban extent med (%)	6 – 20		Urban
urban extent high (%)	25 – 60		Urban
ground truth low	0.05 – 0.3	Which values are used to find ground truth	Urban
ground truth high	0.35 – 0.65		Urban
flood cluster parameter	4 – 8	Determines how many clusters formed for regionalisation	Flood frequency analysis
min AMAX record length (years)	20 – 40	Minimum AMAX record length for consideration in regionalisation	Flood frequency analysis
MAF error threshold	0.1 – 0.4	Max allowable error in MAF calculation	Flood frequency analysis
rain outlier lower	0.8 – 1	Determines what rainfall measures are determined to be outliers	Flood frequency analysis
rain outlier upper	1.4 – 1.7		Flood frequency analysis

Table 5. 1 - Initial ranges set on parameters.

As mentioned, the parameter inputs were created using the "radial" sampling strategy (Saltelli & Annoni, 2010). This strategy is best described as a combined one-at-a-time sensitivity analysis (Song et al., 2015) and a Latin hypercube analysis. Through the parameter space, an initial collection of points is taken. At each point in the space, a single coordinate is varied. As each coordinate is defined by the value of the parameter, this represents the derivative of the parameter from which an elementary effect can be computed. A condition of the strategy involves keeping a minimum distance between the points of comparison in the parameter space, that is Δx_i. When this value is very small, the analysis may produce type I errors, that is assigning an insensitive parameter as sensitive, because small non-linear changes to the output get extrapolated to general model behaviour (Saltelli et al., 2007).

5.3.5. Selection of outputs

It is important to select appropriate metrics when looking at the model results. As there are outputs generated across the entire domain, it is necessary to represent these as singular values per simulation, so that the parameter effects can be interpreted appropriately. Although the model produces results objectively, the conclusions which are drawn from the sensitivity analysis are dependent upon subjective interpretation, and the prior decision of which outputs are considered. Both the total extent of flooding in the domain and the average depth of water might be valid outputs to investigate, but they may report different behaviours, which can be a result of a difference in domain characteristics. If the global flood model is applied in an area with expansive floodplains, we would expect the total flooded extent to represent a better signal for parameter sensitivity. By contrast, if modelling smaller, bounded catchments with narrow valleys, the depth of the waters is a better indication to us of which parameters are sensitive. Hence by considering both outputs we reduce the possibility of committing a type II error; wrongfully assigning a parameter as insensitive (Saltelli et al., 2007), given the effects of topography on sensitivity and flooding inundation (Samela et al., 2017; Shastry & Durand, 2019).

5.3.6. Bootstrapping

Because we are using a relatively small sample, there will be uncertainty in the values which are produced by the sensitivity analysis. We are not certain that the sensitivity indices are converged. To verify convergence, bootstrapping is used (Hailegeorgis & Alfredsen, 2017; Sarrazin et al., 2016). This means the sensitivity indices are recalculated with smaller subsets of the outputs and the variation in the final value is observed. When viewing the elementary effects plots below, this is shown by the boxes which enclose the points. These boxes represent the limits of what other values the statistics of the elementary effects have taken, using a subset. When there is an overlap of the boxes in one of the plots, this implies that the ranking of the parameters has not fully converged, as in some instances, with a slightly smaller sample, one would find that the ranking of the parameters could be changed.

5.4. Results

Monte Carlo simulations were performed at 3 domains: Sacramento, Mississippi and the Po river basins, described in section 5.3. Along with each of these, the total effect of every

simulation has been superimposed on top of each other, so that the recurrence of flooding across each cell can be seen. The following figures represent the sensitivity of the parameters with respect to the extent of flooded area simulated and then also the mean elevation of the targeted downstream floodplain. Hence, they measure the responsiveness of flood hazard to the changes in the parameter values, without comparing the parameter values or outputs to local data.

Further to this we have an estimation of the uncertainty of our sensitivity indices through the boxes enclosing the points. While the mean and the standard deviation are represented by points on the graph, these are enclosed by boxes. These show the potential variability in the sensitivity indices due to bootstrapping, that is taking a smaller subset of the elementary effects and recalculating the indices. If the indices don't change, this indicates a convergence of the sensitivity estimate.

Interpretation of results

The Morris method is applied through use of the SAFE toolbox (Pianosi *et al.*, 2015). This toolbox produces graphs which plot the mean and the standard deviation of the elementary effects along the x and y axis respectively. The higher the value along the x-axis, the more influential the parameter. The y-axis indicates that the parameter is more interactive or non-linear. The magnitude of the x-axis is related to the magnitude of the output considered. It tells us the expected average change in the output value by moving from the minimum to the maximum value in the parameter range. The y-axis presents the expected standard deviation of this change in output. As the model evaluations occur across the parameter space, the non-linearities and interactions which occur between parameters have an effect on the model response. This is captured in the standard deviation of the elementary effects and is shown here.

Also noticeable is the major changes in magnitudes of the x and y-axes across the locations. These values are normalised across the various of parameter values, but maintain the magnitude of the values measured. Thus, when examining flooded area, the perspective is that of thousands of square kilometres being potentially flooded. In the case of assessing the models for flood elevation variance, the magnitude is measured in metres. As these numbers are normalised across the parameter ranges to make them comparable, they do not represent

the literal amount of flooding. Hence it should be remembered that the value of these plots lies in their relative magnitudes rather than the exact values cited.

5.4.1. Po results

5.4.1.1. Parameter sensitivity with respect to flooded area extent

In Figure 5.4 we see the sensitivity of the model to the Po basin domain, with respect to flooded area. We see that the channel Manning's *n* is the most important parameter, which confirms the usual findings of hydraulic sensitivity analyses that it is highly important. Following this is the reach accumulation threshold parameter, which determines how the domain is disaggregated to be modelled by the underlying hydrodynamic model. This is a parameter which is specific to the Bristol GFM and is required due to the modelling limitations of LISFLOOD-FP, as the solver is unable to model the entire domain, as it is unknown how discharge values will propagate all the way downstream in a single domain. Floodplain friction (fpfric) shows a small, additive level of importance to the flooding, and from bootstrapping we see that its sensitivity is well defined, as ranking from the 3rd to the 6th most important parameter. We then have 3 parameters which determine the size of the river channel, which as described in chapter 3 is a particular scheme of parameterisation unique to the Bristol GFM, necessary for a scheme that explicitly models the channel and floodplains and are related to the reach accumulation threshold. They are the minimum/maximum reach length parameters. It is worth noting again that the depth of the rivers is found by using the Manning's equation by moving the depth term to the left-hand side of the equation, given that every other term is known (see section 3.2). This therefore requires an estimation of the discharge of Q when the banks are full. However, this bankfull discharge can be modified by the urban defence parameters, which tries to simulate urban defences by modifying the depth of the channel further to prevent flooding at low return period discharges. As such these parameters are quite interconnected and it is not a surprise to see them together, as they are likely highly interactive.

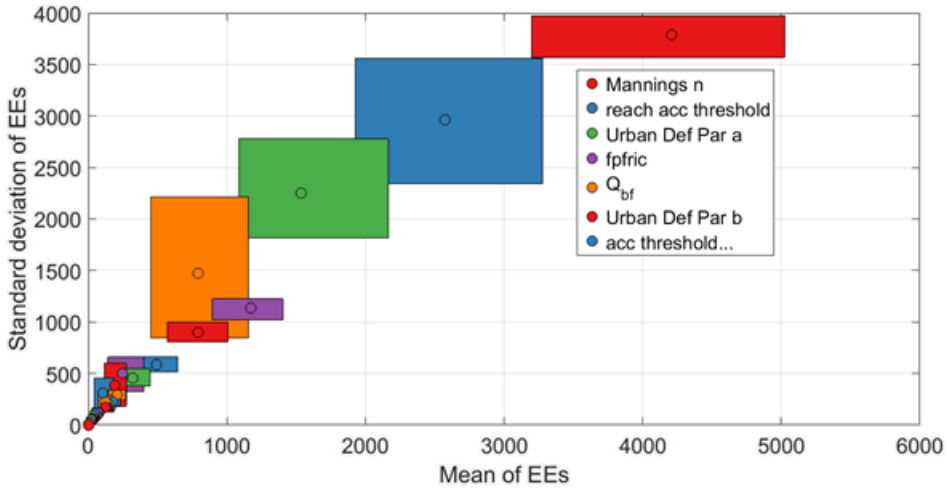


Figure 5. 4 - Elementary effect plot of the Po river, using total extent as the output metric. 7 parameters are identified as important here.

From the above figure alone, we would confidently screen the parameters to include only the first 7. Due to the continuous nature of the calculated elementary effects, the decision of what constitutes an important parameter is subjective. In this case, the decision is based on rejecting any parameter which has a maximum sensitivity of less than 10% of the average sensitivity of the most sensitive parameter. However, in other cases it also comes down to how the parameters are clustered and coming to a reasoned distinction between sensitive and insensitive parameters. With reference to the bootstrapping, we are unable to verify the ranking of the most important parameters.

5.4.1.2. Empirical probability maps of Po basin

With reference to figures 5.5 and 5.1, we can examine how the sensitivity is propagated through the model domain and why. Figure 5.5 presents an image of the Po basin, and each cell represents for what proportion of our simulations that cell is flooded. If the cell has a value of 1 that means within every simulation this cell floods, whereas for 0 it means that it never floods for any parameters set at the 100-year return period. Therefore, values of around 0.5 equal the areas of maximum uncertainty and most sensitive to the parameter values in space. This implies a spatial component to the uncertainty of inundation, confirming other studies (Savage et al., 2016).

Most noticeable is that the main floodplain of the model appears at first as quite sparse, with many of the cells not flooding at all. This suggests that there is not enough water being routed onto the floodplain, although we see that in many of the tributaries there are areas which flood continuously.

Also notable is that all the tributaries are regularly flooding as well. As each cell is a kilometre wide, all of these will be represented at the sub-grid scale, hence while the tributaries are out of bank it is hard to understand at what scale. However, it should not be a surprise that the tributaries would all experience local flooding, given that they are all the recipients of their corresponding 1-in-100 year flow.

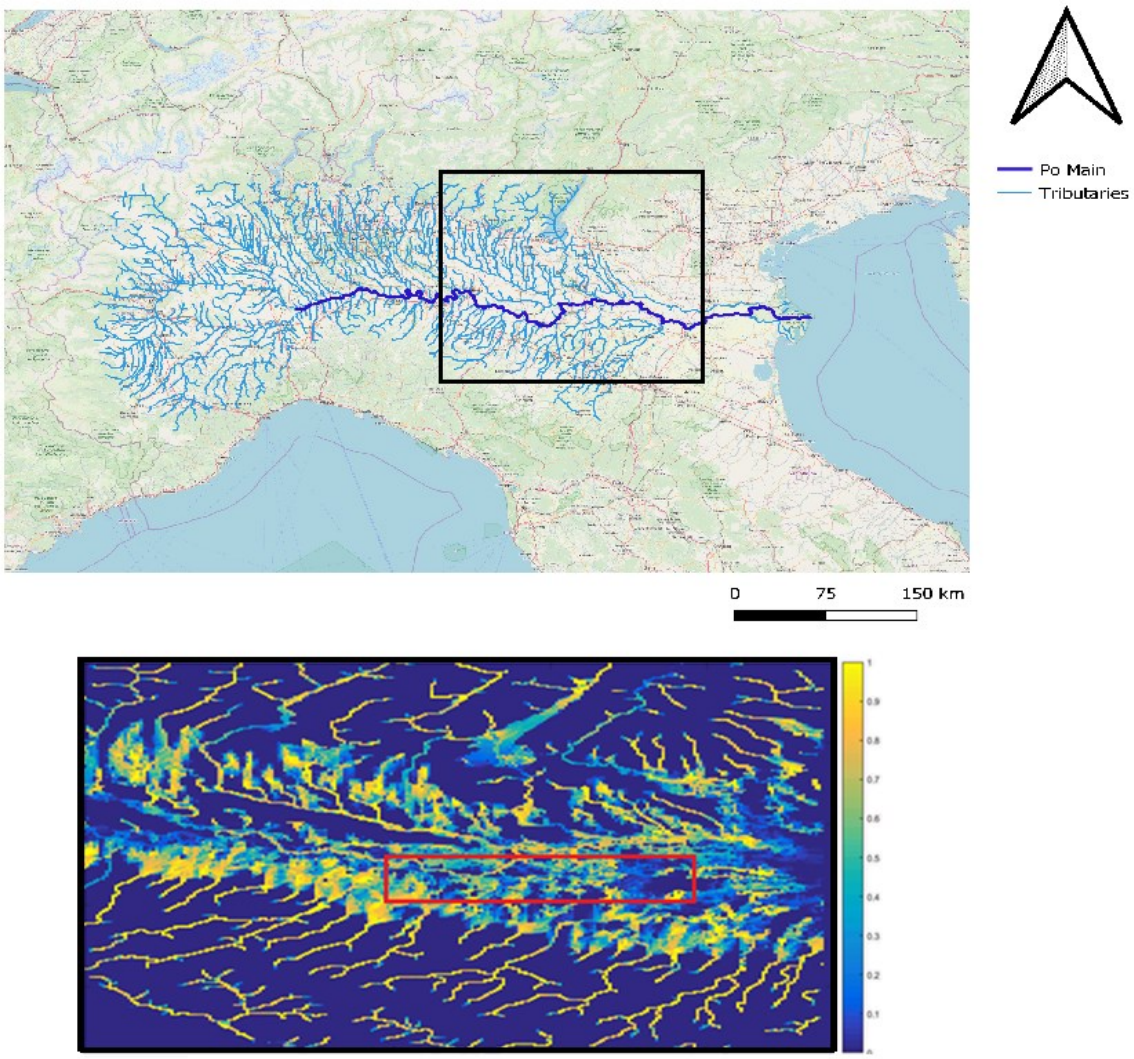


Figure 5. 5 - Empirical 2D CDF of the Po river basin, with area considered for depth analysis highlighted in red, with the larger Po basin highlighting the extent of the modelled domain. Also visible is the Garda Lake in the top part of the domain.

5.4.1.3. Parameter sensitivity with respect to floodplain elevation

Figure 5.6 shows model sensitivities when using elevation as an output metric. Elevation is chosen instead of depth, to a given an absolute value, as use of depth would mean that the

sensitivity would be relative to the variability of the DEM in the local domain. This area must be constrained as many of the smaller tributaries have different elevations due to the topography. Therefore, it only makes sense to consider the elevation in an area where we could reasonably expect the elevation to be changing because of the flooding event. As such, the floodplain is focussed on for producing this metric and record the variance in the average elevation of flooded cells within this zone, as shown in figure 5.5.

We see that by focussing on the elevation, the parameters reach acc thresh and Manning's *n* are again the most sensitive parameters. The Q_{bf} parameter is less important. This is likely since the Q_{bf} is mainly concerned with modifying the bathymetry of the river and hence channel size, while Manning's *n* also has a clear hydrodynamic effect on the distribution of the water, via the velocity of the flood wave. The clear assertion of reach acc thresh as the most important parameter accounting bootstrapping indicates the importance of floodplain decomposition in the Bristol GFM.

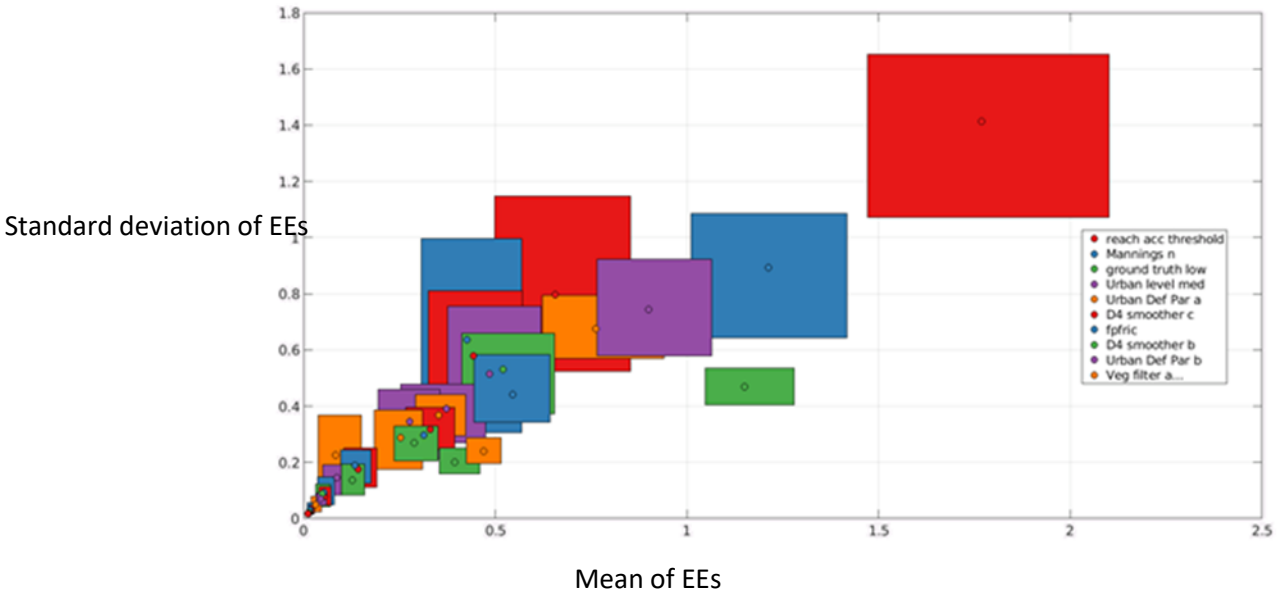


Figure 5. 6 - Elementary effects plot of the Po river basin with respect to elevation of the highlighted area of fig. 5.5. In this case the x and y axis correspond to the change in elevation height anticipated with respect to a change in the corresponding parameter.

5.4.2. Mississippi results

5.4.2.1. Parameter sensitivity with respect to flooded area extent

For the Mississippi, in figure 5.7, we see that there are four parameters that significantly influence model outputs, although there are some other minor parameters which have sensitivity indices which appear large compared with the other sites. Again, we see that Manning's n, bankfull discharge Q_{bf} and the reach accumulation threshold are the most sensitive parameters. There is also a parameter, max reach length, with a large degree of uncertainty, which can also be important. This parameter is linked to the reach accumulation threshold, but rather than being related to an associated dataset (in this case accumulation), it is the upper limit on the maximum reach length which can be modelled individually. It also has a high standard deviation implying that it is interactive with some other parameter or components. However, the size of the box due to bootstrapping is a sign that the actual sensitivity of this parameter has not yet been found – the sensitivity indices have not converged, or its behaviour is extremely non-linear. Given that its position in the overall modelling framework as next to the reach acc thresh, shown here as highly important again, there is clearly some dynamic which is showing highly inconsistent behaviour from the parameter.

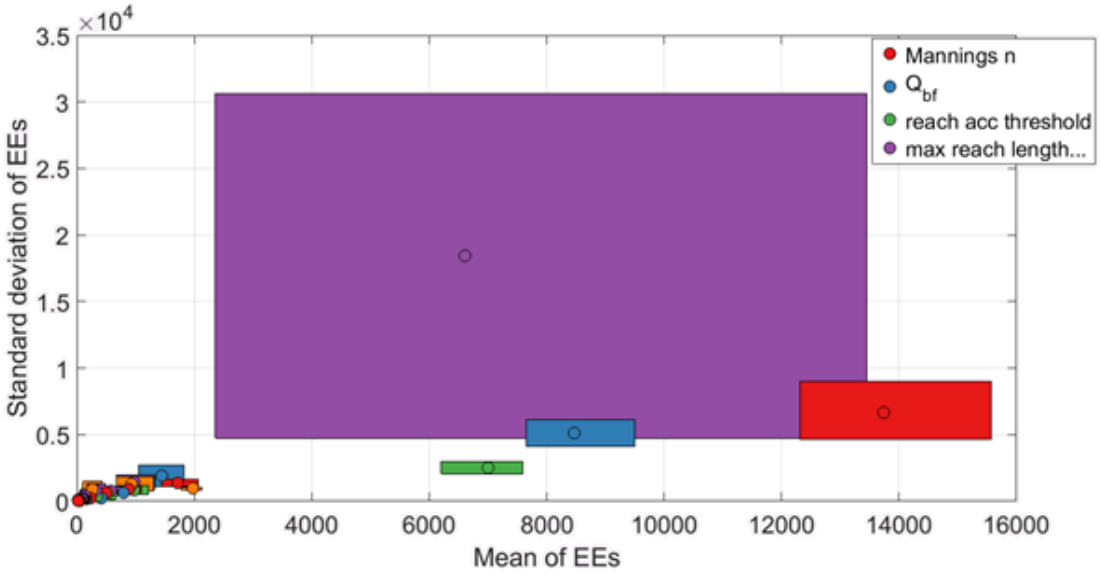


Figure 5. 7 - Elementary effects plot of the Mississippi lower basin with respect to total flooded extent.

5.4.2.2. Empirical probability maps of Mississippi basins

The empirical 2D PDF of figure 5.8 shows that much of the wetted floodplain, which covers an area of approximately 40,000km^2 in our model domain, is highly uncertain, indicating direct dependence on the parameters. Indeed, this indicates that there is a particular parameter combination which would be necessary to correctly represent the true 100-year floodplain hazard on the lower Mississippi basin – no flooding would indicate structural error while constant flooding would indicate equifinality. It is possible that much of this uncertainty is directly relatable to the uncertainty of the max reach length, and perhaps if this uncertainty were constrained then much of the uncertainty on the main floodplain would diminish as well. However, the main point is that the importance of the max reach length is unique to the lower Mississippi in the study, indicating the role of uniqueness of place on the parameter sensitivity within our GFM.

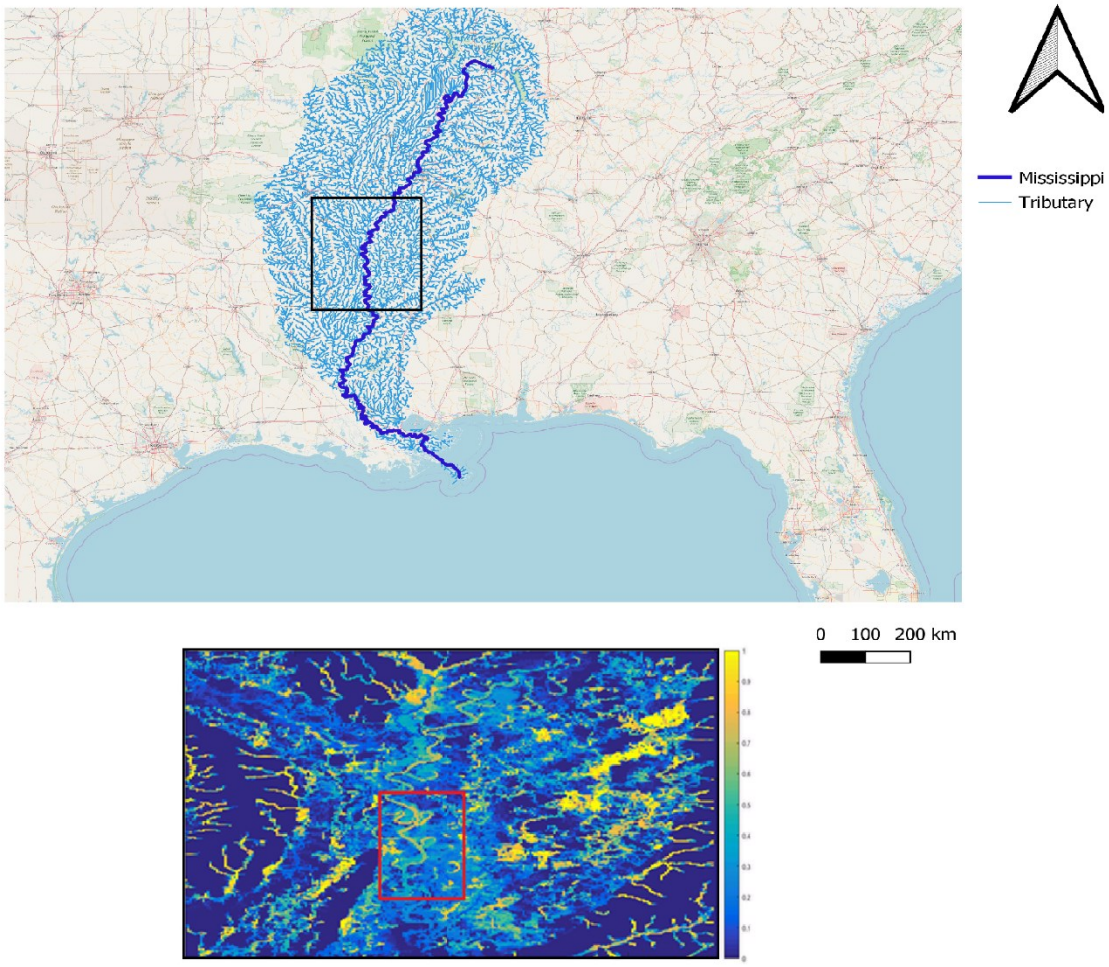


Figure 5. 8 - Empirical 2D PDF of the Mississippi river basin, show with its positioning on a map. Also highlighted is the area chosen for measuring average elevation.

5.4.2.3. Parameter sensitivity with respect to floodplain elevation

With respect to the elevation shown in figure 5.6, along a central part of the Mississippi floodplain, we have the same important parameters, with the same relative levels of uncertainty. Also interesting is that Q_{bf} is a sensitive parameter with respect to elevation whereas at the Po it was not. This is a curious difference and is likely due to the topography of the Mississippi basin compared with the Po. Also notable is the shift of importance from urbanisation in the case of the Po to vegetation in the case of the Mississippi. This is informed by the vegetation and urbanisation datasets which are used at each location, therefore influencing model behaviour through the DEM and channel size respectively. This also makes sense as there are more dense areas of vegetation in Mississippi relative to the Po, whereas

the Po has become one of the most heavily developed lower basins in the world (Montanari, 2012), which is also confirmed by the input datasets introduced in section 5.3.

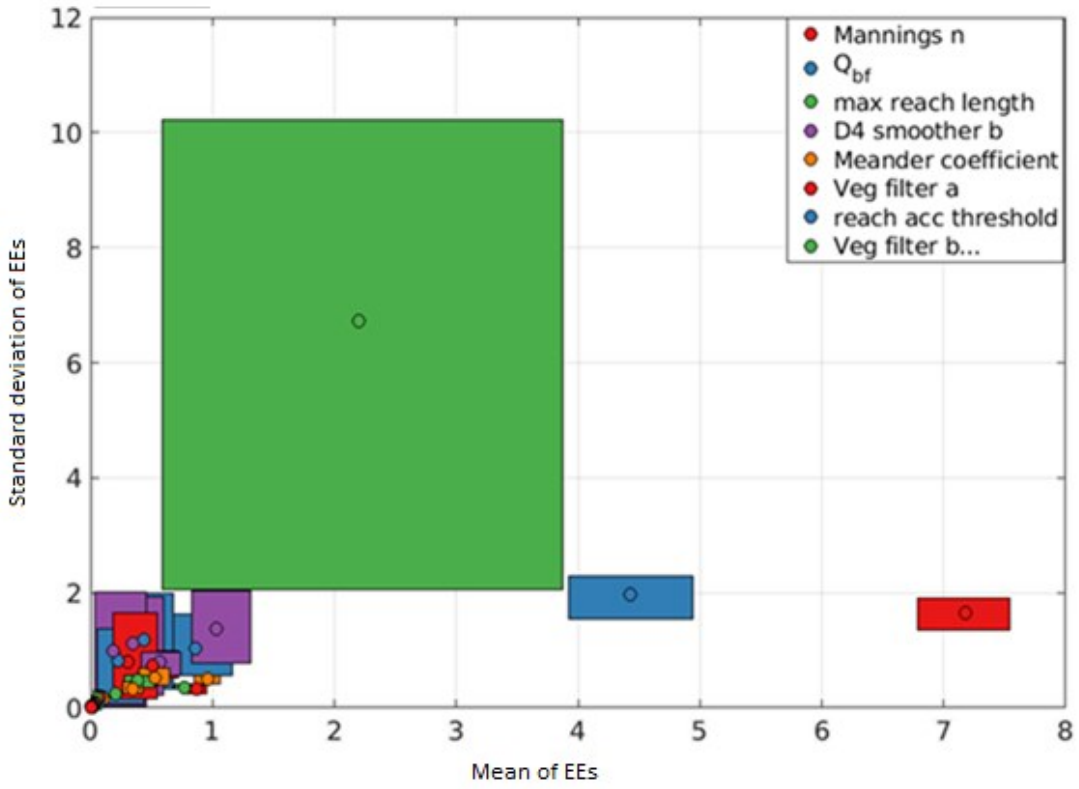


Figure 5. 9 - Elementary effects on the Mississippi lower basin with respect to average elevation.

5.4.3. Sacramento results

5.4.3.1. Parameter sensitivity with respect to flooded area extent

The Sacramento simulations show model runs which are a bit easier to interpret; there is a clear gradation in parameter sensitivity, shown in figure 5.10. There is less interaction between the parameters, as the standard deviations of the elementary effects is relatively low in all cases – indicating fairly linear additive relations between the parameters. Again, we find the 3 most important parameters are Manning's *n*, reach acc thresh and Q_{bf}. Following these we have two further clusters of parameters with lesser significance. This means that for a calibration/ optimisation of the model, one could take 7 or up to 13 parameters in future calibration, and account for virtually all the sources of parameter uncertainty to which the model is sensitive. This number of parameters would be based on what the user subjectively feels is an appropriate threshold of sensitivity.

Among these parameters there is “arid screen rain” which is related to the aridity, indicating that the semi-arid landscape is in some way partially important. This parameter determines whether the channels defined by topographic analysis are likely to be real, based on the amount of rainfall present. The rainfall of the Sacramento domain is such that the threshold condition is occasionally met, meaning that some channels are removed from the modelling domain. As the climate of the Sacramento domain is semi-arid, this makes sense, and highlights one of the flaws of the modelling methodology, that determines the removal of channel networks based on rainfall. This is quite inadequate in terms of the various hydro-morphological processes which produce a river, such as ignoring the effects of geology, and base flow to sustain arid rivers (Renard et al., 2010). In particular, the effect of sub-surface systems which transport large water quantities are not accounted for in this parameterisation.

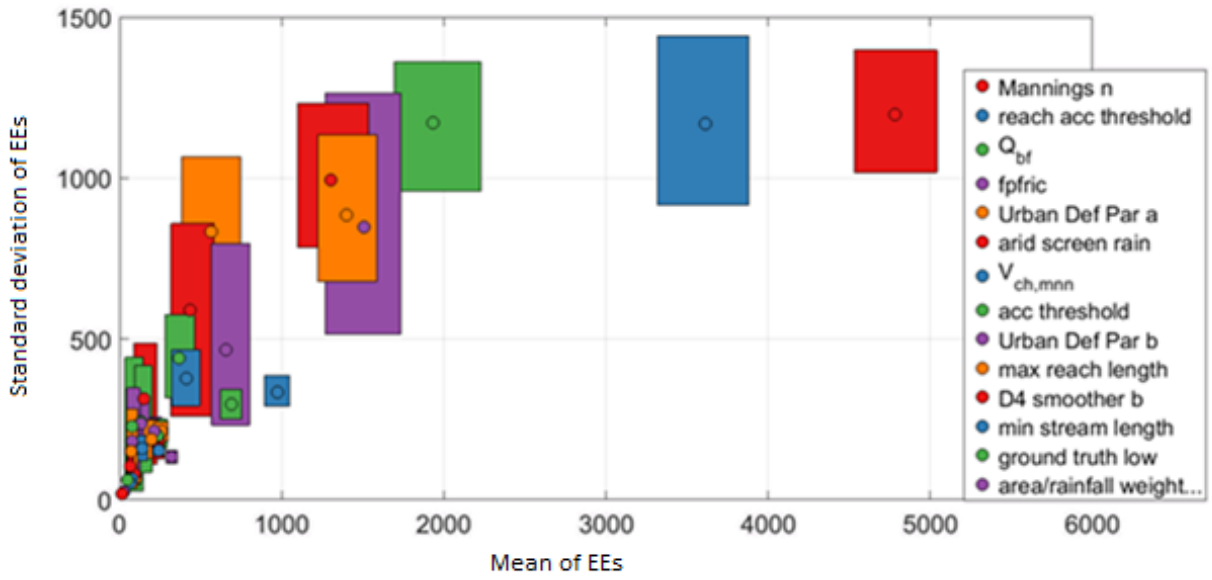


Figure 5. 10 - Elementary effects of the Sacramento river with respect to flooded area.

5.4.3.2. Empirical probability maps of Sacramento basin

The empirical PDF of figure 5.12 shows the variability in flood extent simulated by the GFM, in terms of smaller area flooding in the tributaries and the larger extents on the Sacramento river floodplain. What is notable however is that the floodplain floods more often than some of the rivers on it. This is highly unusual, and implies that there has been some unrealistic computation of the river bathymetry in the model for some parameter sets. Given that the depth of the river is calculated, it could be that there is flooding occurring on the floodplain, but it then re-enters the river further downstream, or the water is sitting on the floodplain. Given that we can see there are many tributaries all flowing into the same floodplain region, this could be a very flood-prone area, meaning that such behaviour might not be impossible, although this is the first time this phenomenon has been noticed in the GFM results. Another possibility is that as a delta, the tributary banks are elevated about the floodplain, not uncommon in deltas, given the effects of sedimentation. But in figure 5.11, this phenomenon is not seen. With reference to figure 5.12, we see that this area is indeed the deltaic region of the Sacramento before it flows out into the Bay Area. The representation of deltas and wetlands is not well understood in the model at the current time, and it is likely that this structural error is due to a lack of adequate process representation with respect to the bathymetry estimation.

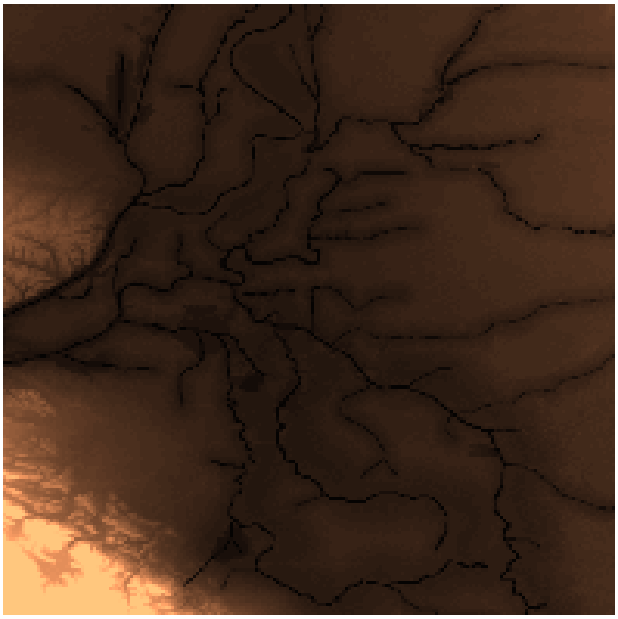


Figure 5. 11 - Zoomed in DEM of Sacramento floodplain, showing deep areas for tributaries. Corresponds to elevation outline (in red) shown in figure 5.12.

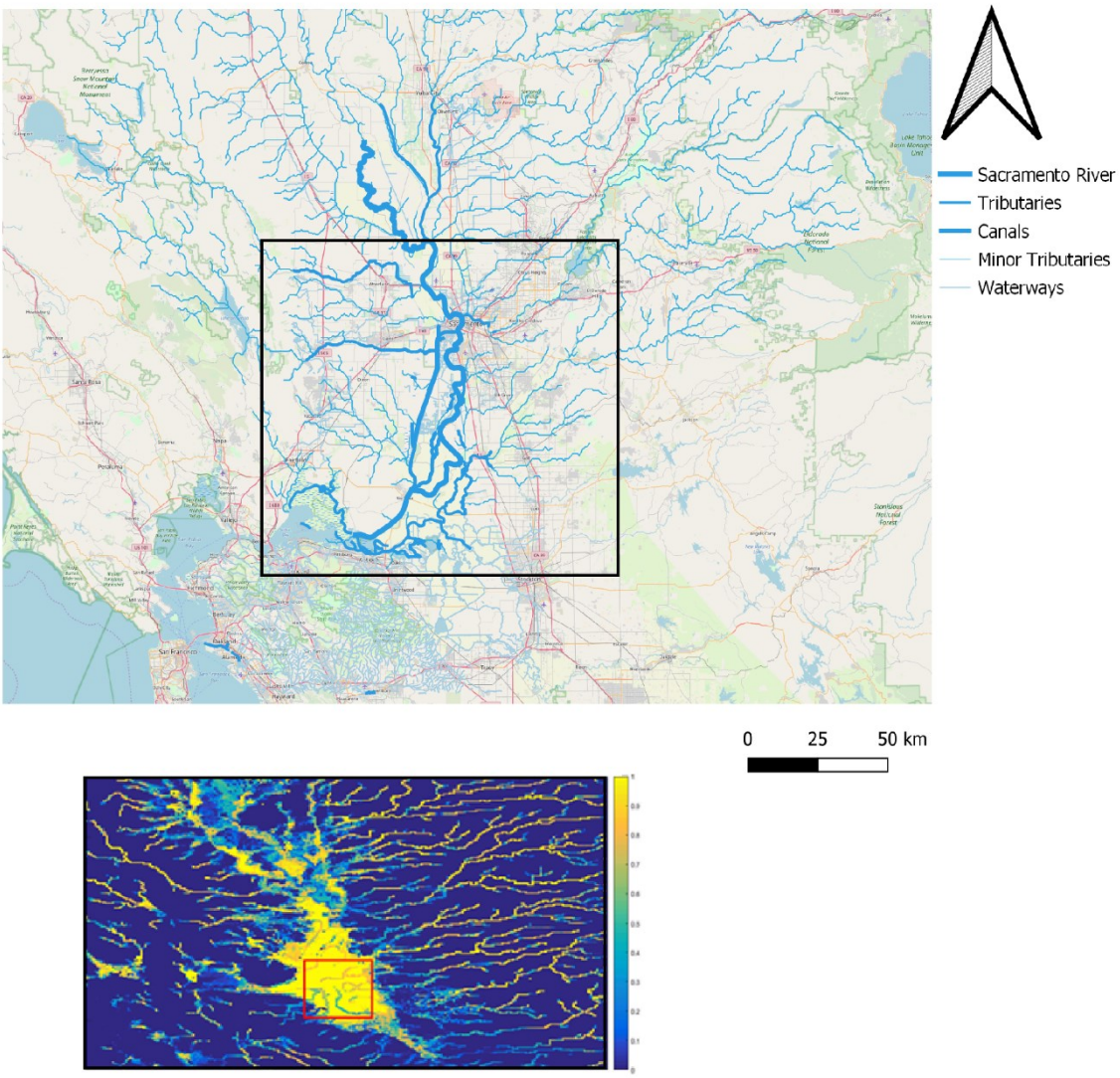


Figure 5. 12- Empirical 2D CDF of the lower Sacramento river, with area considered for depth analysis highlighted in red, with the Sacramento Delta highlighting the extent of the modelled domain.

5.4.3.3. Parameter sensitivity with respect to floodplain elevation

Figure 5.13 indicates there is not a convergence of the parameter sensitivity. When this is compared with the activity seen in figure 5.12 at the delta, it is clear that there is some anomalous activity within the model. Such results provide little information in the way of ranking the parameters; there would have to be a significantly larger number of model runs to converge to a ranking of the parameters, although there is still some value to the plot in terms of the screening. This is due to the effect of the delta on the floodplain area – there is not much uncertainty in any of the cells, except for the river cells. As such, the response of the parameters becomes less clear as many of their effects become interactive in this highly specific scenario. Alongside this it is noticeable that the Q_{bf}, previously a top-ranking

parameter does not make the top 10 when looking at elevation. Given its sole effect on bathymetry, this is difficult to explain, if we assert that the lack of flooding in some river cells is due to an increase in depth. So, either this is the reason, or the river depth at these points is not the main factor. However, without a fuller convergence of the parameters it is not possible to give a definitive answer in this context.

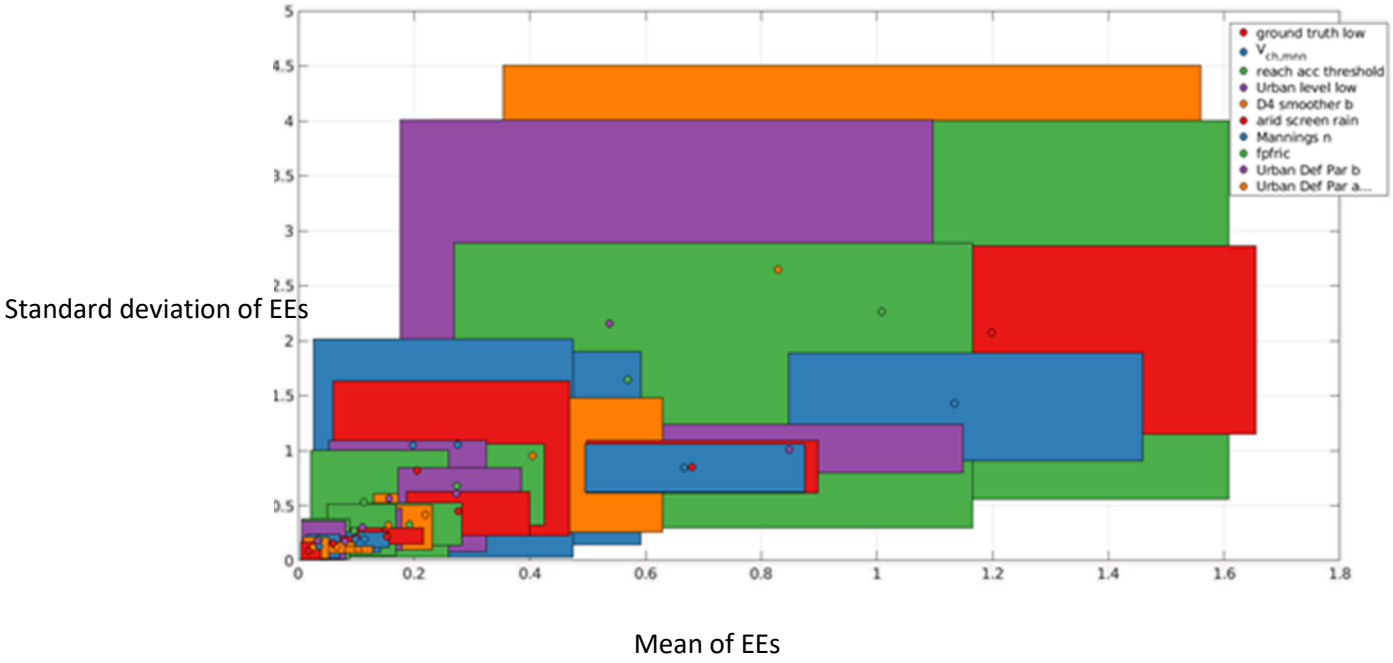


Figure 5. 13 - Elementary effects test of Sacramento river with respect to the average elevation.

From these results, it is clear that there are 3 parameters which are consistently highly important. Were someone to focus on only 3 parameters in every calibration of the global model, the Manning's *n*, reach accumulation threshold and bankfull discharge would be the 3 parameters of choice for a practitioner of the Bristol GFM. Although these parameters wouldn't exist in the same way in other global flood methodologies, there would likely be some analogue parameterisations which we would therefore also expect to be highly important. This can be said with some confidence because the functionality is very likely to be similar across all models which focus on inundation. Channel size, length and friction are all decidedly important components of flood inundation modelling, and that is reflected here in the most sensitive parameters.

5.5. Discussion

What could we expect to learn from the sensitivity analysis, besides the most sensitive parameter set? It can give feedback on whether the model is appropriately parameterised and whether some constraints can be placed on the parameter scheme. Whilst any model can suffer from equifinality, which is where different parameter inputs return the same output, it can often be an endemic feature of more complex models (Beven, 2006; Schoups *et al*, 2008). It is normally indicative of over-parameterisation, where there are several parameters which are performing an equivalent function.

Two of the three most important parameters are both hydraulic. These are Manning's *n* and the bankfull discharge parameter Q_{bf}. Topographic parameters do not show the same level of importance to our output metrics we have chosen. Given what we know about the uncertainties of the DEM and its application to flood modelling (Domeneghetti, 2016), this is slightly surprising. However, on this point it should be made clear that the topographic parameters do not fully capture the uncertainty of the DEM, but rather only aspects of it. Parameters which are commonly not important are the climate parameters. These focus mostly on the arid regions however, which for the Po and Mississippi domain weren't involved. Distinction of climate is instead done by the regionalised flood frequency analysis, which clusters catchments based on Koppen Geiger climate classification. The rFFA parameters themselves aren't very important, despite their use in a theoretically highly important module of the GFM. Perhaps because they are concerned with the data-driven analysis of the method of Smith et al., 2015, this therefore means that they have an indirect and hence secondary effect.

One of the key innovations of the Bristol GFM is the estimation of Q_{BF}, the parameter which is used for the inversion of the Manning's equation. This is a necessary process to allow calculation of the river depth. The Q_{BF} parameter range is in units of years, to represent the return period interval of the bankfull discharge flow. This return period interval is plugged into the rFFA routine, as discussed in section 3.3. While it is generally held that the expected return period interval for temperate rivers is 1.5 years (Williams, 1978), it has been thought that there is still significant natural variation in this estimate, which has recently been categorically (Edwards et al., 2019), lending credence to the wider parameter ranges which we gave the parameter. The end result of this parameter is to affect the total conveyance of

the river reach, via a greater or lesser channel depth. Should the return period be every 5 years, then the relative difference in flow compared to the 100-year event is smaller than for a 1.5-year return period for the Q_{BF}. This excess flow is what is propagated across the floodplain and is therefore what gets represented as a hazard. Therefore, we can see directly how the Q_{BF} parameter is tied to the prediction of hazard in the Bristol GFM.

In the domains analysed, the most important parameter is only occasionally identifiable – it has clearly converged above the rest. Convergence is defined in this case as the likelihood that a parameter's sensitivity has been correctly identified (Sarrazin *et al.*, 2016). However, the ability to rank the other remaining individual parameters is mostly not possible, and insensitive parameters, whilst not converged individually, have converged in groups to be clearly unimportant, although it must be conceded this importance is decided through a subjective judgement. There are three broad classes of parameters which we can identify.

- Evidently insensitive – from the ways that we have looked at the data, and the supposed behaviour that these parameters exhibit, we can leave them at the default value without affecting model outputs in terms for inundation extent or water levels.
- Evidently sensitive – these are parameters which across multiple locations come up as the most important parameters which in a calibration experiment of the model one would want to consider in order to improve the skill.
- Varying sensitivity – these are parameters which vary in their importance across different locations, with respect to the same output metric. What this tells us is that the behaviour of the parameter is not static, but dependent on the input data that is used in each model. As the data is a function of the physical land characteristics, it follows that the parametric behaviour of the model is a function of the geoclimatic characteristics of the given domain. That is to say that for a global model with globally-oriented parameters, in local-scale applications, there will be tendencies to have unique features, which will only relate to a subset of the parameterisations available to the model. Furthermore, as we change the metric of parameter sensitivity, we focus on

different aspects of the modelling domain. This difference is seen by assessing the sensitivity of the model when using elevation as a metric.

5.5.1. Convergence

The question of convergence is a key determinant of success of the sensitivity analysis (Sarrazin *et al.*, 2016). In the analysis performed here, there is irregular convergence of individual parameters. However, we do see across the various analyses of the model that there are a few parameters which characterise behaviour within a given domain. This is seen clearly in the method performed at the Mississippi. The "importance" of these parameters is converged, and similarly, there is a convergence of the other parameters of no sensitivity in the outputs observed.

5.5.2. Standard deviation of elementary effects

The standard deviation of the elementary effects of the parameters are not yet completely understood. Whilst the explanations are limited to either the model being non-linear or the parameters being highly interactive (Saltelli et al., 2007), these could be due to threshold combinations of parameters which cause anomalous estimations of flood hazard. This means that a combination of parameters must have specific values in order to produce outlier results. However, the sampling procedure used here is too sparse to give us anything more than a suggestion of this type of complexity, which is outside the scope of this study. A more intense form of analysis such as VBSA would be required to answer this. However, the plots give indication that the parameters in the top right corner are much more interactive than the parameters in the bottom left corner. This isn't necessarily true, because the standard deviation value uses the distribution of EE's; if the parameter is highly sensitive, it is much likelier to have a higher standard deviation value, because the values of the distribution will be of a greater magnitude.

5.5.3. Average elevation as model output

We find that where there is a reasonable level of convergence, the elementary effects due to elevation support the elementary effects due to total area. This however isn't the case with the Sacramento. We tried multiple floodplain areas to try and find some sensitivity results and in each case the parameter rankings and sensitivities varied wildly, with levels of uncertainty shown in the above example. It is quite unclear why this should happen for the

elevation here when the total area extent appears stable and conforms to the results seen for the Mississippi and the Po.

5.5.4. Hypotheses of important parameters

We found that many of the parameters conformed to our expectation of what would make them sensitive; the most obvious being Manning's *n*, confirmed as important in previous flood modelling studies (Hall et al., 2005; Pappenberger et al., 2008). Another parameter which has consistent importance is the reach accumulation threshold. This parameter determines how the model defines reaches, based on the upstream accumulation calculated by HydroSHEDS. An increasing number of rivers consequently increases the potential flooding occurring within a given domain, as the corresponding number of boundary conditions increases as well.

It was also anticipated that Q_{bf} would be a highly important parameter, in its regulation of the channel depths, and characterising the size of the channel relative to return period flows, by defining at what discharge a channel exceeds bankfull stage. However, in the case of the elevations we also see some evidence that it is not so important for the elevation, as the Po results show.

5.5.5. The role of input data in determining parameter sensitivity through domain types

An area of future interest from this work can be to ask why the parameters vary in sensitivity across location, and whether there is some systematic structure to this, which we expect there to be. From the perspective of a parameter, the only thing they "see" is the input they interact with. Hence a changing input with the same parameter would lead to a changing output sensitivity. So, if parameter sensitivity is changing when this input changes, we could hypothesise that the change in parametric sensitivity is a function of the changing input; which in this case are the input data layers.

The analysis gives an indication that this is the case. For instance, consider the arid rain screen parameter. In one of the domains it becomes highly important, because the inputs become commensurate with the threshold of the parameter, causing it to function and remove some of the river channels if there is not the threshold amount of measured annual precipitation within the domain. Hence, if this threshold is varied across the measured rainfall in the domain, then each individual model evaluation will have different channels in it, leading to different flooded areas, so the parameter is highly important.

5.5.6. The apparent unimportance of the regionalised flood frequency analysis

Of the parameters which have shown themselves to be unimportant, perhaps the most unexpected is the response of the regionalised flood frequency analysis parameters. A trivial expectation of the analysis is that surely calculation of the peak flow is important. However, instead here the calculation of the peak flow as it is represented by the parameters has little importance. It seems likely that the uncertainties of this part of the model simply cannot be represented by the parameters.

5.6. Conclusion

A global sensitivity analysis of the Bristol GFM was conducted at multiple locations, to understand the behaviour of the model distinct from the characteristics of the locations chosen to model. The main aim was to screen the parameters of the model as either important or not important, to reduce the complexity when seeking further improvement and study on model uncertainties. The approach used the Morris method, but parameters were often sampled too sparsely to rank them individually. However, we were able to identify subsets of parameters which are important, and have begun to unravel some of the behaviours of the model. Our main conclusion is that there are three parameters which are consistently highly important. Were someone to focus on only three parameters in every calibration of the global model, the Manning's *n*, reach accumulation threshold and bankfull discharge would be the three parameters that we recommend they focus on, for this specific model. Despite the dependence of these parameters being specific to the Bristol GFM in this study, there is likely a transferability of their importance. This is because similar processes (friction, conveyance capacity or river, and how the model domain is broken into reaches) are represented in all GFM's and will therefore have to deal with.

According to our analysis, there are some parameters which are always important, and these aren't just concerned with the hydraulic aspect of the model. Instead, they also include parameters which relate how the inputs are modified to become the domains. This is significant for other GFM's which although may have different methodologies and cascades associated with them, will nonetheless still be reliant on global datasets which need to be modified prior to use. Similarly, there are also parameters which have been identified as always unimportant, which can be ignored in further analysis, to facilitate a more intensive

sensitivity analysis (e.g. VBSA) or model parameter calibration. The importance of individual parameters is not static across location but varies and it is thought this is due to the characteristics of a given location. This can be understood without requiring the total convergence of the individual rankings, as it is still possible to identify subsets of parameters which are qualitatively more important.

From this work we see that the following investigations could be made:

- With some of the unimportant parameters identified, their values can be set at default and with a smaller set of parameters, a deeper uncertainty-based analysis assessing the model's performance could instead be undertaken.
- To make sure that some of these unimportant parameters haven't been misidentified, an analysis where the most important parameters have their ranges severely constrained/set to a singular value could see the other parameters doing other things. A key example in this case is that the peak Q has not been represented here as a sensitive parameter, despite the expectation that it would. This implies that the method of prediction is deterministic rather than parameterised.
- An investigation into which are the characteristics which determine what makes some of the parameters sensitive only some of the time.

5.7. Post-script

This chapter represents a searching phase of the thesis, with the general aim of acquiring a better understanding of the model. Prior to the work done here, although there have been examples of the Bristol GFM's performance, there has been no understanding of how or why that might be the case specifically with respect to the model parameters. Further to this, there is very little work like this being done on any other GFM. Hence, some of the difficulty of this work has been in trying to deal with the complexity of the Bristol GFM while simultaneously not having any other examples to compare with. This is one of the reasons why the raw variance was preferential to using performance metrics, which will be considered in the next chapter.

While the work done so far doesn't help us to understand any aspect of the model precisely, it allows for the correct perspective to be taken on model parameterisation. Having found that most of the parameters do not matter, or have highly circumstantial effects, the actual mechanisms which determine output variance can become a much closer object of study. This will be seen in the second results chapter. Alongside these aspirations, it might also become clearer why some of our expectations regarding the most important parameters have been confounded. This chapter concludes a firm base on which to establish some more rigorous analyses of the modelling framework.

6. Research Chapter 2 - Application of a Global Scale Model to a Gauged Local Basin: Assessment and Uncertainty Analysis

6.1. Introduction

In the previous chapter, a basic sensitivity analysis of the Bristol GFM was undertaken and, given this, some general conclusions regarding the behaviour of the model and some specific parameter sensitivities were made. With this knowledge, we now have a general direction towards better understanding the model's response to its parametrisation. However, although we are now able to say qualitatively which parameters are sensitive, there is still the task of *quantifying* these behaviours, according to some more specific measures of performance outside of raw variance.

In this chapter, we focus on assessing the Bristol GFM at a single location – the Po river basin. With the knowledge obtained from the previous chapter, there is a smaller pool of parameters on which to focus. With this, and some data pertaining to the behaviour of the Po river, an assessment about the skill of the GFM when applied to the river Po will follow. Further to this, the behaviour of each model evaluation is used to weight the parameter values, thus returning posterior parameter distributions. This means that model parameterisations with uncertainty, that are appropriate for this specific domain, will emerge. This will be able to help inform what is the potential performance of the GFM when using local data.

The application and assessment of the model at a specific location creates new questions about the local relevance for a global model, which can be partially answered here, in the context of the Po river. While it is often held up as a virtue to have finer resolution DEMs for inundation models (Schumann et al., 2014), it has been demonstrated that there are diminishing returns for model skill as the resolution increases (Savage et al., 2016, Jarihani et al., 2015). As such, it is to be seen if the coarse representation of the domain through the various remotely-sensed data sets will be able to simulate flooding at the Po basin, with the added assumption that there are fundamentally no errors with the DEM product used. In the case that it is not, then the incorporation of local information may be leveraged, to assist. This also has interesting implications for the GFM community as it begins to ask to what degree should a GFM be considered global, if it begins to use local datasets.

Much of the work in this chapter falls largely within the conceptual framework which is used by the GLUE methodology. By this, the use of Monte Carlo sampling, prior and posterior parameter distributions and the informal definition of conditional probabilities are all used therein. This chapter was conducted as a part of the visiting placement segment of the PhD, at the University of Bologna with Professor Montanari.

Despite all GFMs sharing a principle objective of simulating hazard, it has been demonstrated that there is inconsistency across the various methodologies, given then same location and return period (Bernhofen et al., 2018; Trigg et al., 2016). However, there are also specific examples of the models performing well (Wing et al., 2017; Yamazaki et al., 2014) and showing improvements in performance with additional information (Bernhofen et al., 2018; Wing et al., 2018). As a consequence, it has been proposed that there should be a way to systematically compare results of various flood models, with standardised performance metrics and experimental setup (Hoch & Trigg, 2019). This approach will be able to distinguish how the GFMs vary, allowing the use of ensembles or probability weighted mapping (Dankers et al., 2014; Di Baldassarre et al., 2009). It will also be important to consider what the constituent structural errors are within each model, therefore allowing for appropriate model identification and setup given the state of the problem. This has been done in catchment hydrology with the use of FUSE (Clark et al., 2008) and MOPEX (Duan et al., 2006) for example. To answer these questions for a GFM, one needs to think about how they are applied at the global scale how it can be validated therein.

6.1.1. How to validate a GFM

Given the emphasis on model comparison in the literature to date, we propose that a crucial area of study for GFM's is an assessment of their individual uncertainty (Beven et al., 2011; Hall & Anderson, 2002). This should also proceed in a systematic way (Montanari, 2007), with clear definitions of the various sources of uncertainty which can lead to model error (Götzinger & Bárdossy, 2008). Flood events are by definition hydrological extremes, leading to sparse datasets and high uncertainty (Levy & Hall, 2005). How to parse uncertainty has been cause for much debate in hydrology (Beven & Binley, 2014; Montanari, 2005; Stedinger et al., 2008), particularly surrounding the phenomenon of equifinality (Beven, 2006). Without commenting on its truth value, it seems the case that the informal methodologies proposed by Beven et al (Keith Beven & Binley, 1992; Keith Beven & Freer, 2001) are applicable for

GFM's because their usual application to ungauged basins means it is difficult to formally establish parameter priors. In general, GFM's are only valuable where detailed local studies do not yet exist, which is in fact most of the planet.

6.1.2. Can a GFM be locally relevant?

The concept of uniqueness of place presents a challenge to GFM's in that they must maintain a consistent methodology while appropriately representing local-scale processes (H. V. Gupta et al., 2014). Although, there are examples of a continental scale inundation model being applied at the local scale and successfully modelling inundations, indicating that GFMs have potential in this regard (Komi et al., 2017; Siqueira et al., 2018; Fleischmann et al., 2019) and despite these promising advances, there are not yet studies which evaluate model behaviour in depth at a given location, accounting for parametric and structural uncertainties.

In this context we will examine the Bristol GFM. Using a Monte Carlo approach, we will assess performance and evaluate whether it is "behavioural", drawing on approaches which apply a GLUE-style methodology (Franks et al., 1998; Aronica et al., 2002; Blazkova and Beven, 2009; Liu et al., 2009). By linking the parameter values through the use of conditional probabilities to the performance, we will be able to map out the uncertainties of the parameters, and the implications of this for performance of the Bristol GFM at the local scale (Ammann et al, 2018; Aronica et al., 2002). A key conclusion of this method of study will be to establish whether what constitutes a skilful model output will be equivalent to a realistic model output, with good process representation (Beven, 2007; Beven, 2000).

6.2. Methodology

While in the first results chapter there was a sensitivity analysis which focussed on a screening approach, in this chapter the work is focussed more on identification of specific parameter behaviours. This means using a reduced number of parameters and also a denser sampling scheme. This produces a range of inputs and outputs within the modelling framework, to understand how uncertainties are propagated through the modelling chain.

As there is an emphasis on assessing the performance of the model, an MC sampling methodology as described in chapters 2 and 4 is used, using non-informative priors. Based on the performance of each simulation, the given parameter value has an associated weighting,

which provides a posterior distribution. In this way, we can understand which settings present us with realistic predictions of flood hazard.

6.2.1. Case Study – Po river basin

The subject of our study will be the Po river basin. The Po basin is a major European basin and the largest in Italy. There are approximately 15M inhabitants of the basin, and the region produces 40% of Italy's GDP. Hence the risk of flooding can have very serious consequences. In this sense, it represents exactly the sort of basin which needs protection globally, although there are very few of this size which are monitored to the same degree. Most basins of this size and importance globally are still ungauged or have restrictive access to gauge data and are poorly understood – such as the Chao Phraya basin of Thailand.

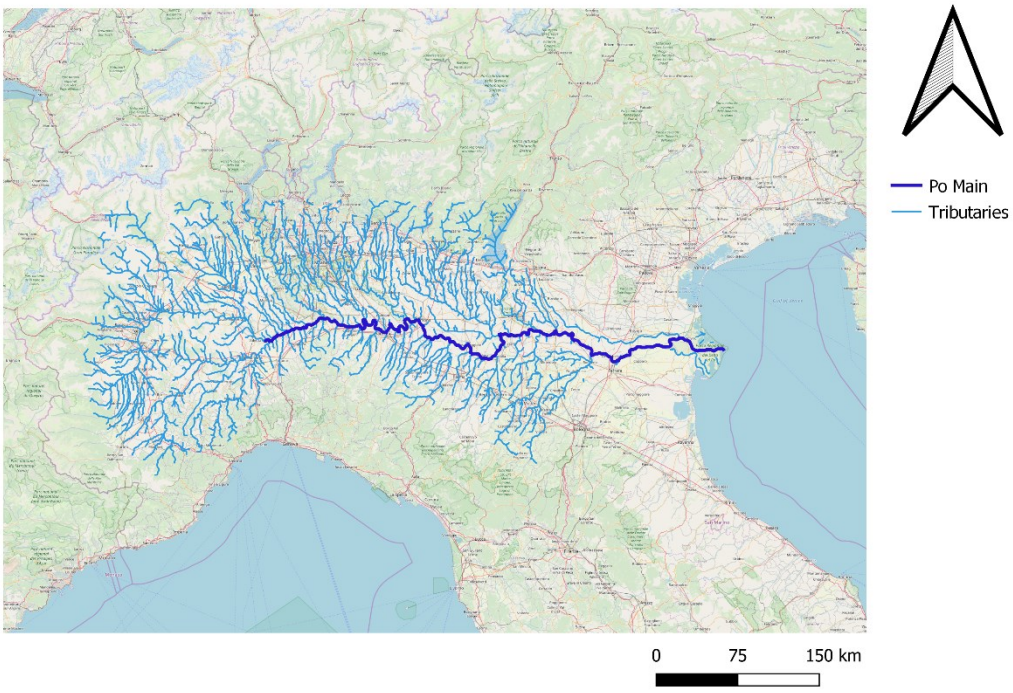


Figure 6. 1 - Po basin and major tributaries. Produced using dataset of Andreadis et al., 2013.

The Po river is a seasonal river and the majority of its precipitation falls in the mountainous regions which are at the limits of the basin. This means that seasonal snowmelts form the Apennines and the Alpine catchment and are the main drivers of peak flows in the Po during the springtime melting period. There are 4 gauging stations along the main stem of the river for which we have data to derive ratings and annual maxima time series. There have also been

studies conducted by the Po river basin authority to model the flood hazard for the 100-year flow, which we can use to assess the performance of the Bristol GFM (AdBPo, 2012).

6.2.2. Parameter selection and model setup

From the screening conducted in the prior chapter, the selection process for the parameters has mostly been done, as we selected the highest ranked parameters from the previous chapter, which we are confident dictate the behaviour of the model. With a parameter space which is dimensionally smaller, it can be sampled more densely in order to capture further information about model behaviour. The basic number of model simulations used is 5,000 model runs. While more would be ideal, the number is limited by the amount of time taken for a single model run and the computer resources available, as discussed in chapter 3.

In the previous chapter, sensitivities were recorded by sampling just the output variability. This gave the raw sensitivity of the model to the parameter inputs, unbiased by any expectation of performance. In this chapter, we use in-situ observations to provide a benchmark of ideal model behaviour, given that there is now an understanding of the output variance. This means our focus has shifted away from just assessing the raw variance of the model, and instead how well the model can reproduce some basic attributes of the Po system.

Table 6.1 displays the prior parameter ranges chosen for each parameter; these were selected to be as broad as feasible to account for maximal uncertainty, while attempting to avoid the use of non-behavioural parameter values:

Parameter	Range	Description
Urban defence par a	0.01 – 0.095	Defence parameters which modify the depth
Urban defence par b	0.5 – 0.85	of channel according to urbanisation (see table 1)
Bankfull discharge (Q_{bf})	1 – 5	Return interval of bankfull discharge
Manning's n	0.015 – 0.06	Determines roughness of riverbed.
Reach acc threshold	0.5 – 1.6	Decides length of river reach per sub-model
Floodplain friction (fpfric)	0.025 – 0.15	Determines roughness of floodplain

Accumulation threshold	50 - 1000	Threshold of upstream area to determine a channel exists.

Table 6. 1 - Parameters selected and their ranges for the analysis of the Bristol GFM.

As discussed in chapter 5, the justifications for Manning's and bankfull discharge could be derived from literature. The bankfull discharge was selected to represent a range of potential rivers, but still mainly confined to temperate or semi-arid areas (Castro & Jackson, 2001; Williams, 1978). Given our knowledge of the Po's climate, this seemed sensible. Manning's *n* was selected to cover a large range, stopping at roughness estimates which are at the extreme ends for this river system given its morphology. This is also true for the floodplain friction (Arcement & Schneider, 1989; G. Di Baldassarre & Montanari, 2009).

In the case of the urban defence and the reach acc thresh parameters, as these are specific to the model, it is difficult to give any physical justification for their selection of number. Recall that the bankfull discharge is modified based on the degree of urbanisation, represented by the ISA datasets, meaning that the bankfull discharge in urban environments is calculated as:

$$Q_{BF,urb} = a \times ISA^b$$

Where *a* and *b* are the representative urban defence parameters. Given the ad hoc form of this parameterisation, the selection of ranges is necessarily arbitrary and their values were chosen by varying the default parameter value approximately ±60-80%, to make them identifiable without becoming overly sensitive. The reach accumulation threshold's upper limit was extended further, as it was found in preliminary study that with the upper limit below 1 there was a missing peak in the results.

6.2.3. Sampling

The sampling approach here was a Latin hypercube all at a time uniform sampling across the parameter ranges (Wainwright et al., 2014). The prior parameter distributions were all assumed as uniform, because in all cases, there were no stronger assumptions about the nature of the parameter distributions which could be made. This is particularly so for the parameters that are unique to the model because there is no prior data on which to base assumptions of a particular distribution, normal or otherwise.

Arguably, the Manning's coefficients have a tendency towards a central moment in their range. In many cases it is clear that a roughness of coefficient of n = 0.03 is suitable, and is

often used in many studies as the default – it is in fact the default value of the Bristol GFM (see chapter 3). However, if assigned a prior non-uniform distribution, the sensitivity of Manning's with respect to other parameters would be masked because of the prior expectation of its higher sensitivity. Therefore, it is better to allow a uniform distribution sampling and then observe a posterior distribution of values, to clearly demonstrate the response of the model.

6.2.4. Definition / choice of performance metrics

To better understand the behaviour of the model and identify the parameters, the assessment is conducted against multiple metrics, each representing a different 'dimension' of the model outputs. The output distributions will be evaluated in terms of comparison of the flooded area, water surface elevation, bed profile elevation, and peak discharge.

Maximum Flooded area

The first performance metrics used will examine the performance of the GFM maximum flood extent simulations for the 1 in 100 year hazard against a local flood hazard map. The local information available for this will be the flood hazard maps of the Po river basin authority shown in figure 6.2. As discussed in section 4.2.4.1, this hazard map is the best available benchmarking data. While the optimised GFM set up would ideally produce a theoretically perfect critical success index of 1 (see below), this would not mean that the global flood model has perfectly represented the actual 100-year flood event. While the local map comes with some information about the depth of the flooded hazard, this is not uniformly distributed throughout the modelled domain. Therefore, we limit ourselves to benchmarking against the flooded extent of this hazard map.

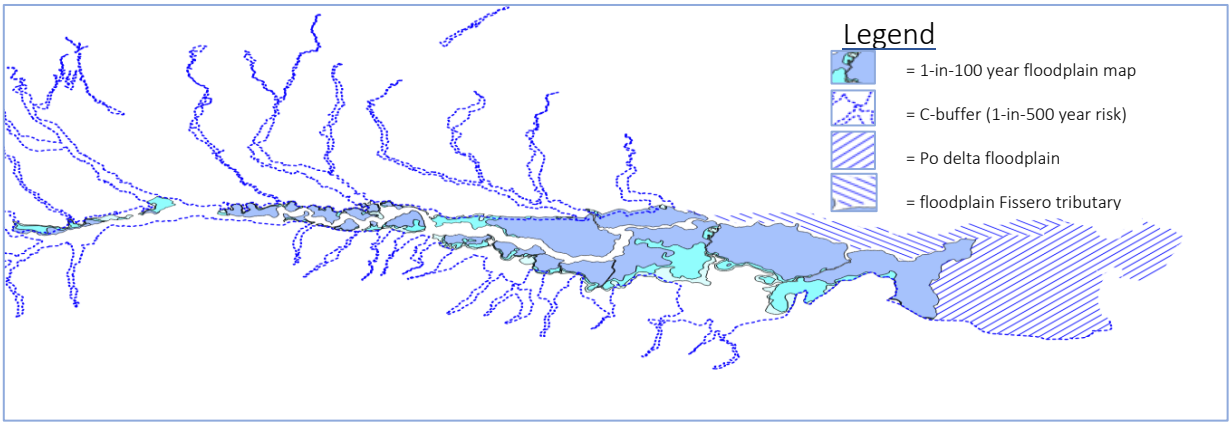


Figure 6. 2 - Po river basin authority flood map: provides outlines of 4 different areas of hazard (AdBPo, 2012).

Our approach to assessing the performance of the flood inundation model based on area extent follows on from the approach of Aronica et al., 2002. This is appropriate given we are making an area comparison of 1-in-100 year flood hazard maps. In practice this is done by a cell by cell comparison, which records true positives, false negatives and false positives per model evaluation, against the local hazard map. Each model's performance is evaluated using the critical success index (CSI) (Bates and de Roo, 2000):

$$CSI_i = \frac{A_{obs} \cap A_{mod,i}}{A_{obs} \cup A_{mod,i}}$$

Where CSI_i is the total fraction guessed correct, A_{obs} is the observed flooded area and A_{mod} is the modelled flooded area.

Equation 1 can be expanded to:

$$CSI_i = \frac{\sum_j p_{i,j}^{D_1M_1}}{\sum_j p_{i,j}^{D_1M_1} + \sum_j p_{i,j}^{D_1M_0} + \sum_j p_{i,j}^{D_0M_1}}$$

Where $P_{i,j}$ is pixel "j" with the model 'i', and D_xM_y represents the classification for that given cell. Hence the CSI is a measure of how well the model is able to reproduce the flooding of the observation. It is a binary classification, where the value of each cell is either 1 or 0 for the model and observation. A perfect model has a CSI value of 1, but the model is penalised for incorrectly estimating the state of any of the cells, which lowers the value towards 0. How these states translate into each of the terms of the above equation is displayed in table 6.2.

Cell status	**Flooded in model**	**Not flooded in model**
Flooded in observation	D_1M_1	D_1M_0
Not flooded in observation	D_1M_0	D_0M_0

Table 6. 2 - Classification of each cell based on whether flooding is recorded for both the model and the observations.

Elevation method

To understand the model in detail, it is not enough to evaluate the model against just the total flooded area. Larger valley filling events exhibit less signal, as the flood extent becomes

insensitive to changes in simulated water levels at the edge of the flood. Instead, the elevation of the water surface can produce a stronger signal of model behaviour for more constrained areas of flooding (Tourian et al., 2017). Hence, we are also comparing our model to the gauges present at the Po river, shown in table 6.3:

Station	***Latitude (°)***	***Longitude (°)***	***Elevation (m)***	***Avg. Velocity (m/s)***	***Avg. Q(m^3/s)***
Piacenza	45.06	9.70	42.37	0.61	933
Cremona	43.15	9.99	29.03	0.82	1075
Borgoforte	45.02	10.75	14.05	0.84	1313
Sermide	44.89	3.48	9.50	0.44	1378

Table 6. 3 - Gauging stations along Po for which we have data.

Within the modelled domain there are four gauging stations, Piacenza to Sermide and we compare the global model simulations of the 100-year water surface elevation to the 100-year elevation (E_{100}) at the gauge site. We derive the E_{100} values of the gauges with the following method:

1. Deriving the Q_{100} value.
    a. Derive annual maxima discharge (AMAX) data from 20-year daily time series. These are fitted using a generalised extreme value (GEV) curve, with a least-squares optimiser in MATLAB.
    b. Use the GEV curve to estimate the Q_{100} value. The GEV modelling tools can also produce 95% CI if required.
2. Deriving the E_{100} value.
    a. The same data set has ~20 years of daily data of both depth and elevation.
    b. A rating curve for the purposes of this study is fitted using a power law model of the form $E_{100} = a*Q_{100}^b + c$. These parameters were estimated using again the least squares optimiser approach.
3. Defining the range of acceptable E_{100} values.

a. In each case, the goodness-of-fit from the power curve and GEV allows us to extract the confidence intervals derived from the uncertainty of the fits to our local data. This uncertainty can be propagated through to our models with sampling if necessary.

The elevation likelihood is calculated differently to the area extent likelihoods. Given that the higher RMSE values constitute a worse performance, the weights of the model must reflect that. For a given model 'i', the RMSE is calculated from the elevation error at each of the gauging stations:

$$RMSE_{elev,i} = \left(\left(E_{GS,j} - E_{i,j}\right)^2\right)^{1/2}$$

Where $E_{GS,j}$ is the elevation at gauging station 'j', and $E_{i,j}$ is the elevation in model i at point j, corresponding to the gauging station location.

Bathymetry profile

Another aspect of the model is the calculation of river bathymetry, which cannot be remotely sensed. Many of our parameters are linked to the calculation of river depth, specifically Q_{bf}, Manning's n and (indirectly) the urban defence parameters to calculate the channel depth. Hence it is interesting to investigate whether the most accurate models also satisfy this aspect of the domain. Channel bed is often not considered during model evaluation although it is certainly important in terms of realism.

We have a river profile data set which gives the elevation of the bottom of the riverbed, located with a downstream distance measure. This topographical survey means that we know the bed elevation at select points within the model. By cross-referencing the latitude and longitude of these points with the cells in our model, we can see whether the river bed has been adequately estimated. There are many associated uncertainties with this approach, given the point source local data and the coarse global representation. It is not clear how the continuous actual bed profile should compare with the discretised form used by the model with assumptions made about its shape and monotonicity.

However, we can propose a sensible way to do this by considering how the model calculates the bed elevation. To estimate the bed elevation, the model first calculates the channel

depths using an inverted form of the Manning's equation. This modelling process is outlined in section 3.2.

Error calculation

A pointwise comparison of elevation is conducted along the Po river domain, using 23 points including the four gauging stations. Selection of points aimed to select channel sections with simple bathymetry. By mostly avoiding meandering sections and confluences, complex bathymetries which will be bias against the model are not accounted for. We already know the limitations of the model in that the channels are assumed rectangular, so there is little point in further penalising the model for an impossible task of representing a meandering section.

Using the error vector, the RMSE of each model's bed elevation can be calculated. A major drawback of this assessment is the constraint of a monotonic riverbed in the global model, something which is not necessary in the local data. The RMSE is calculated by comparing the absolute elevation of points along the river bed:

$$RMSE_{prof,i} = \left(\left(E_{bed,j} - E_{i,j} \right)^2 \right)^{1/2}$$

Where $E_{bed,j}$ is the elevation of the local bed at point j and $E_{i,j}$ is the elevation of the bed in model i at point j.

Q_{100} estimation likelihood

Given that in the likelihood approach for the gauging station elevations the Q_{100} values are already present, it makes sense to use them, and compare the flow rates generated by the global model with the Q_{100} values at the gauging stations.

Comparison is more difficult in this case. The peak flow is only defined at the boundary conditions. However, the location of the boundary conditions is one of the functions of the model – specifically the reach acc thresh parameter determines the positions of these boundary conditions. Hence the comparison of the Q at the boundary conditions (Q_{BC}) with the Q of the gauging stations (Q_{GS}) becomes less meaningful the further apart they get. Therefore, we also consider the distance between the boundary conditions and gauges before deciding if a comparison can be made. This is necessary, as we only want to assess the models when the Q can be adequately surmised. Should the Q of the boundary condition be located far away from any of the gauging stations, it is likely that it will be poor performing and non-

behavioural. This is likely to happen in the simulations where the boundary conditions are sparsely populated within the model.

We look at the performance by calculating the RMSE of each model, like so:

$$RMSE_{Q,i} = \left(\left(Q_{GS,j} - Q_{i,BCj} \right)^2 \right)^{1/2}$$

6.3.4. Conversion of error values into likelihoods

Of the four performance metrics described above, we can use these to provide weighting for each of the parameter values, thus leading to posterior parameter distributions. The weighting of the conditional probabilities (CP's) is derived such that the best performing models have the largest weight. In the case of the critical success index this means the weight can be directly proportional, but in the case of using RMSE this weight must be inversely proportional.

Therefore, we derive the following CP weights for the parameters:

$$CP_{area,i} = \frac{CSI_i}{\sum_{j=1}^{N} CSI_i}$$

$$CP_{elev,i} = \frac{RMSE_{elev,i}^{-1}}{\sum RMSE_{elev}^{-1}}$$

$$CP_{prof} = \frac{RMSE_{prof}^{-1}}{\sum RMSE_{prof}^{-1}}$$

$$CP_Q = \frac{RMSE_Q^{-1}}{\sum RMSE_Q^{-1}}$$

The summation of likelihood weights L is 1, which makes it suitable for calculating the posterior distributions through the CDFs of these values, which are linked to the parameter ranges.

6.2.5. Calculation of the posterior PDF's required some smoothing.

From the above likelihoods, both CDF's and PDF's are calculated. While the CDF's are easy to plot, it was found that due to the proximity of some of the likelihood values, the resultant PDF surfaces – which is essentially a derivative of the CDF function – were extremely noisy. This effect can be seen in figure 6.3. Hence some noise reduction techniques were used to retain the response shape of the PDF's, but this gave a smooth response curve.

The techniques to reduce the noise of the CDFs and PDFs were interpolation and moving window averaging. Once the CDF's were plotted, the curves were interpolated across 100 points. This was necessary because using every point on the CDF returned very noisy PDFs, as shown in the PDF. Therefore, it was considered more effective to use interpolation. Although the PDF evidently points towards the CP of a lower value being used, it is not an appropriate posterior PDF for future use. Using interpolants is partially effective, but to reach the final PDF surface required, a moving window average was employed, to approximate the final smoothed derivative of the CDF's, giving the posterior parameter distributions.

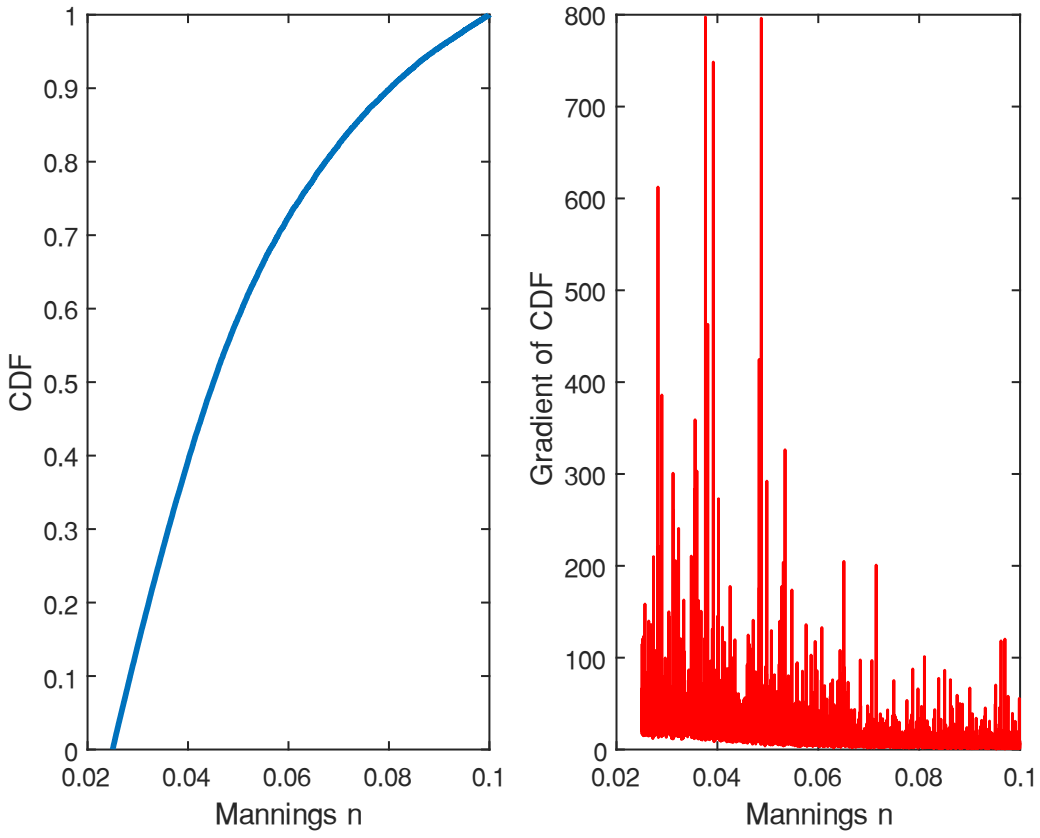


Figure 6. 3 - Calculation of parameter PDF using the actual gradients of the modelled CDF

6.2.6. Uncertainty of observation measurements

Uncertainty of the local data must be considered as well. Particularly given that we are not studying the model and its ability to recreate an event, there are uncertainties around what is a representative 100-year event, as the records we have available are much shorter than 100 years. This is addressed in the use of prediction intervals to cover our local data with 'uncertainty envelopes' (Montanari, 2007).

A key problem with comparisons of hazard is that as both the observed and the predicted values are statistical events, neither can be observations. As the map provides a 100-year return period, we are performing a model benchmarking. However, we are treating the total extent as an observation. As such, we do not consider the effects of observational uncertainty for area.

Derivation of observation elevations is to be calculated from rating curves. It was hoped that uncertainty bounds could then be derived from these, following other methods which have attempted to quantity flow uncertainty at the Po (Di Baldassarre & Montanari, 2009). However, the ratings themselves do not represent true observations at the Po, as the values acquired are daily averages of a record with a much finer timestep. As such, we are not able to quantify the uncertainty in this case as the actual observed structure of the data is not available. While it is clear that the discharge uncertainty should be quantified in future analyses of the model (Coxon *et al.*, 2015), here we must proceed without it.

There are several uncertainties associated with the profile comparison. The exact positioning of each point cannot be fully ascertained. However, this uncertainty is offset by the greater uncertainty of how much the bed elevation could vary with a $1km^2$ global model tile. Also uncertain is whether the errors we encounter will be the result of the model's requirement for monotonicity, or simply due to poor model skill and the poor assumptions of channel geometry.

The Q_{100} estimates are derived from the daily data, which is converted to AMAX and then computed from a GEV distribution. Hence, they have the same epistemic problems as with the elevation data, in that the daily values are averaged from shorter intervals of time and must be treated as deterministic to be useful.

6.3. Preliminary Study

Although the reasoning and evidence for the setup of the model described above is defensible, it was found that initial runs of the model, using the previously mentioned parameters and their ranges, actually produced very poorly performing models, across the entire parameter space. Some of the results of these poor preliminary analyses will be shown here, to convey what led to some changes in the initial setup of the model, but which ultimately leads to a better analysis.

Initial runs of the above setup suffered from substantial underprediction. Figure 6.4 shows the entire domain of the Po river, recording the frequency of flooding in each cell. Cells denoted by yellow have a flood recurrence 100% of the time, whereas dark blue cells never

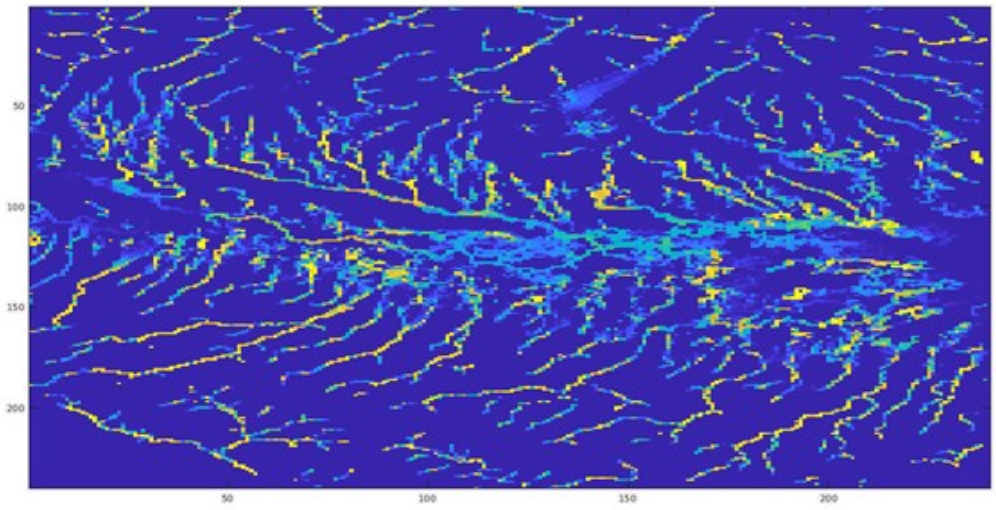


Figure 6. 4 - Empirical 2D PDF of preliminary runs of GFM at Po basin - shows significant underprediction. Cells are colourized to represent a range of values from 0 to 1. Cells which are repeatedly inundated across all 5,000 model runs are equal to 1 and are represented here as yellow.

flood. While many of the tributaries are well represented, the floodplain shows very little evidence of any recurrent flooding, which seems non-behavioural. This conclusion is reinforced by the calculation of the CSI scores. The maximum CSI score in this batch of models is 0.4, while the average is closer to 0.1. This acutely poor performance is clearly due to some other structural reasons, implying that more information should be incorporated into the model setup.

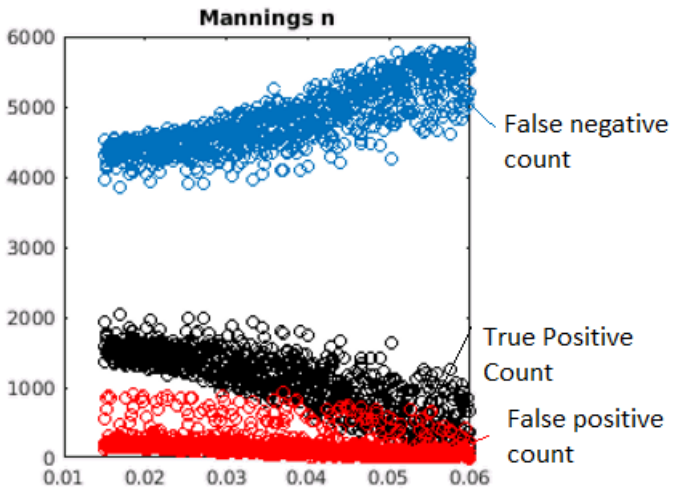


Figure 6. 5 - Plot of cells counts for true positive, false positive and false negative with respect to Manning's n values

6.3.1. Incorporating new information

To establish the cause of this error, the parameters were plotted with respect to the individual components that make of the CSI score – that is, the true positive, false positive and false negative counts. This plot is shown in figure 6.5 for Manning's *n*, the most identifiable of the parameters in the setup. Across the entire space of Manning's *n* it is demonstrated that the reason for poor scores is unequivocally due to the number of false negatives, regardless of the parameter value.

The deduction from this evidence is that the model is structurally underpredicting the extent of flooding. In discussion, it was suggested that the reason for this underprediction in flood extent was potentially due to an underprediction in the peak flow. When checking the boundary condition inputs for the model, it was found that the boundary conditions used for each model evaluation for the Po mainstem were of the magnitude of 6,000 to 7,000 m^3/s. Instead, the peak flow of the Po river for a 100-year return period has an expected value of greater than 10,000 m^3/s (Montanari, 2012). This discrepancy suggests that the peak flow is the reason for the model underprediction, and in a smaller sub-experiment, the expected peak flows given local gauge observations were substituted into a model with randomly selected parameter values, and the overall effect on the floodplain was compared. In Figure 6.6, the effect of this discharge error is clear, with figure 6.7 showing much better representation of the floodplain.

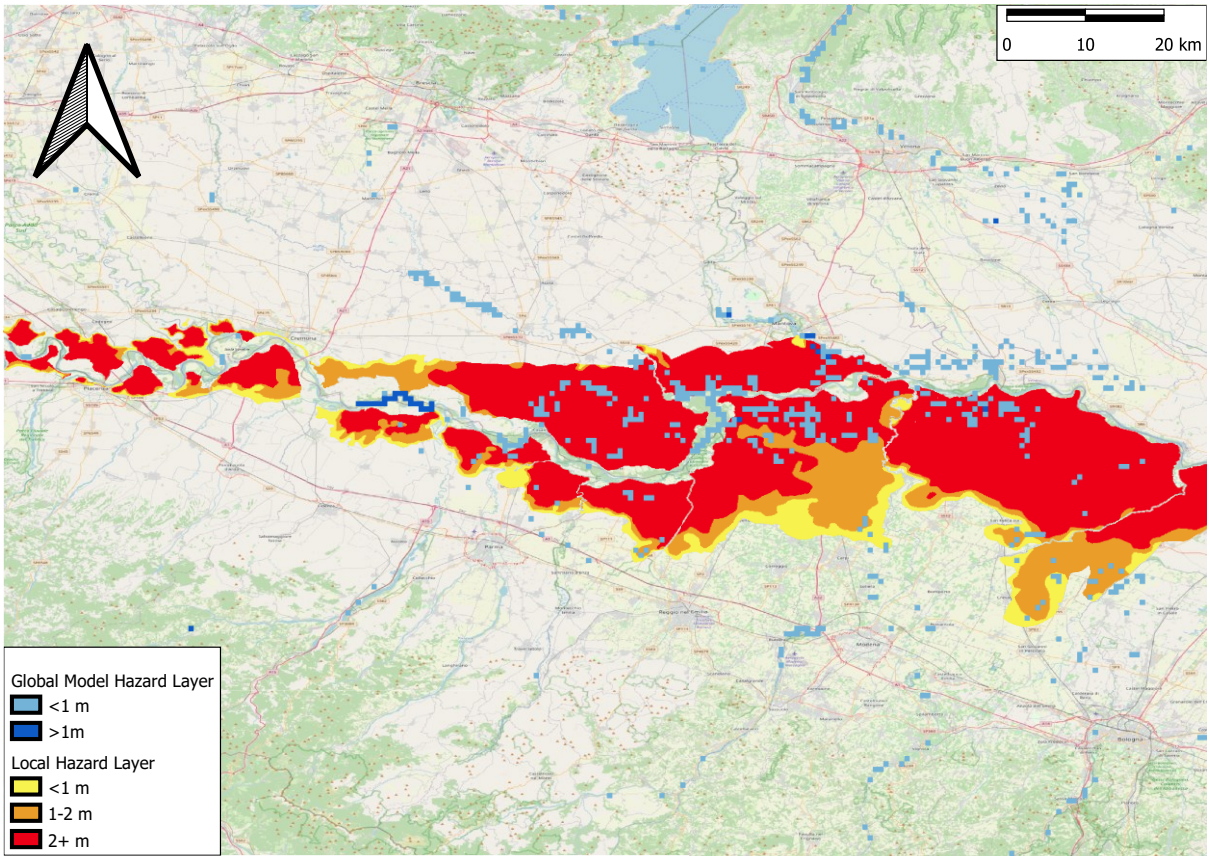


Figure 6. 6 - Comparison of model and observed flows - lower Q estimate.

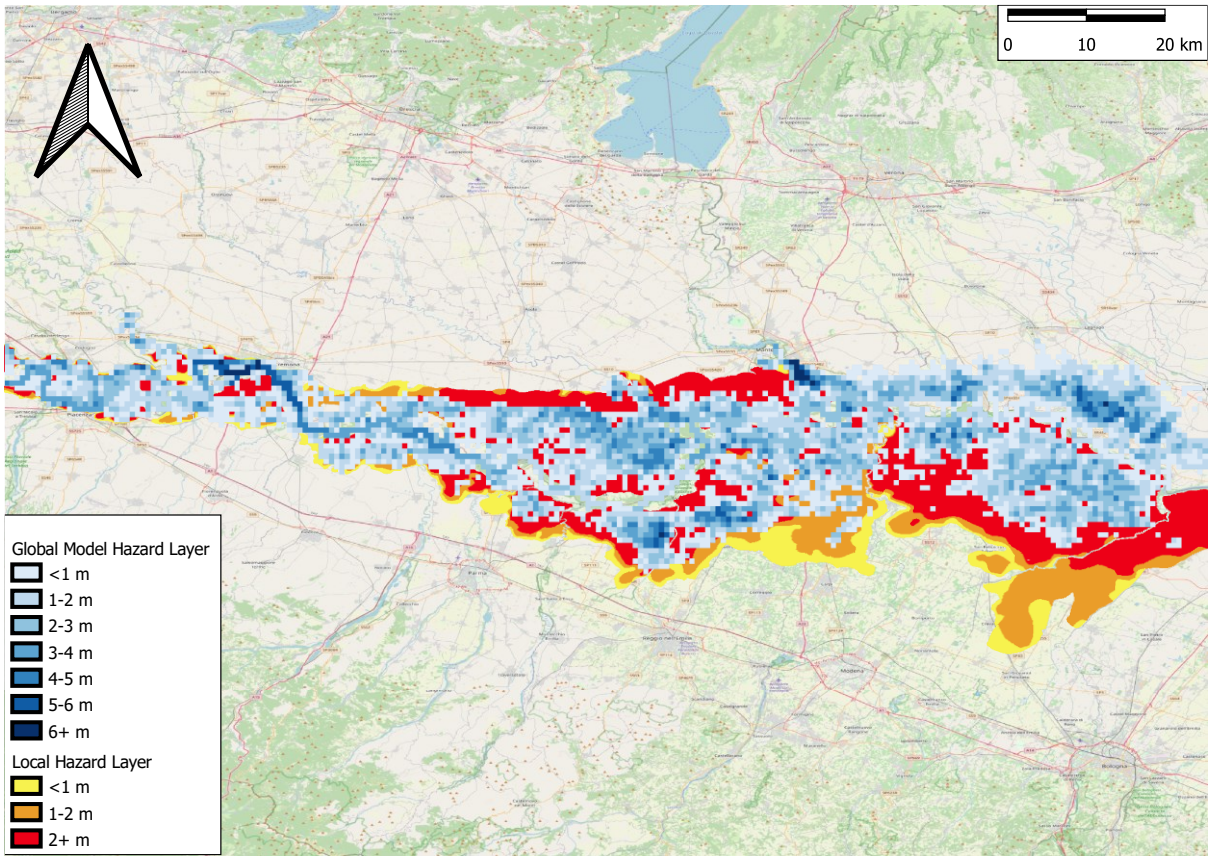


Figure 6. 7 - Comparison of modelled and estimate flows - upper Q estimate.

It is important to assess what the reason for this poor performance is. In the original screening study, none of the parameters related to the regionalised flood frequency analysis indicated any sensitivity at all. The implication of this is that the calculation of the peak flow is determined structurally and cannot be accounted for by the model parameters. Given the amount of uncertainty in peak flow studies normally (Di Baldassarre & Montanari, 2009; Zhao et al., 2017), this clearly represents an oversight in terms of the GFM methodology; however, this is not surprising given the original modelling framework wasn't developed with uncertainty analysis in mind. This uncertainty is critical and requires incorporation into the analysis before further work can be conducted.

6.3.2. Present updated setup for model analysis

To represent this new fundamental uncertainty into the analysis, another parameter is introduced: the peak Q_{100} multiplier. This parameter will act as a coefficient to the peak Q_{100} values calculated by the model's rFFA module. This will act as a proxy for the uncertainty of the peak flow, as there is currently not a method for quantifying it explicitly, although a

method for doing so is developed in the following chapter. The coefficients will be chosen to represent the range of potential values the Q_{100} could be, assuming that the current estimate is the current lowest value.

Besides this change, the decision was taken to remove the accumulation threshold as a parameter, which determines the smallest reach to be modelled in the GFM based on upstream accumulated area. Returning to figure 6.3, it is clear that the tributaries themselves do not struggle to have their peak Q_{100} estimated. Furthermore, the sensitivity of the accumulation threshold has no significance in this setup, as the validation data we have available does not extend to the tributaries. As such, the preliminary study indicates that a posterior distribution can be immediately placed on this parameter, such that only the accumulation of the Po river is counted. In other words, we tell the GFM to only simulate the Po hazard and not the tributaries. There is little point in evaluating the model on the tributaries when there is no information to compare against because these areas will appear to be over predictions by the model when calculating CSI. Hence the updated parameter set is the following:

Parameter	Range	Description
Urban defence par a	0.01 – 0.095	Defence parameters which modify the depth
Urban defence par b	0.5 – 0.85	of channel according to urbanisation (see table 1)
Bankfull discharge (Q_{bf})	1 – 5	Return interval of bankfull discharge
Manning's n	0.015 – 0.06	Determines roughness of riverbed.
Reach acc threshold	0.5 – 1.6	Decides length of river reach per sub-model
Floodplain friction (fpfric)	0.025 – 0.15	Determines roughness of floodplain
Q_{100} multiplier	1 - 4	Increases peak flow at boundary conditions

Table 6. 4 - Updated parameters and their ranges following preliminary analysis.

6.4. Results

The following sections cover the results for the updated parameter set described above, with the posterior parameter distributions calculated with respect to each likelihood.

6.4.1. Area-based CP

Simulations which attempt to recreate the floodplain of the local validation data generally perform well. Figure 6.8 shows the CSI score of each model, comparing to the data of figure 6.2. There is an upper limit of a score of 0.7 for the critical success index. This is a good performance, and one which approaches that found in literature by other studies of this GFM (Sampson et al., 2015).

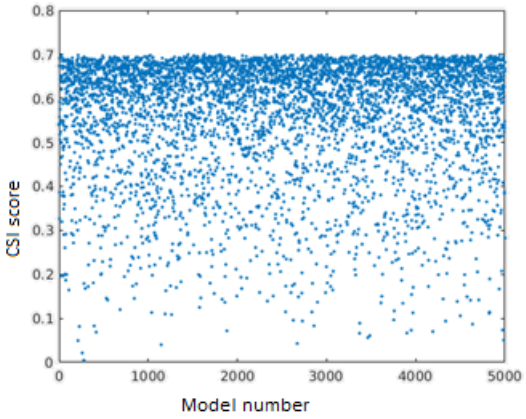


Table 6. 5 - CSI score for each model, derived using updated parameter setup and ranges.

We overlaid the flood map of each model iteration to produce an empirical 2D PDF of the Po basin. Seen in figure 6.8 this shows the frequency which each cell is flooded over each model iteration, with the colours ranging from 0 to 1. We see the central area is well represented, although the model struggles to flood upstream and downstream quite so often. There is also a collection of small areas within the floodplain which consistently fail to flood every time, which can be noted by comparison of this plot to figure 6.2. The reason for this is likely to be related to the quality of the DEM, which can have some rather high errors in it, as a result of urbanisation, vegetation cover, or just poor measurement (Yamazaki et al., 2017).

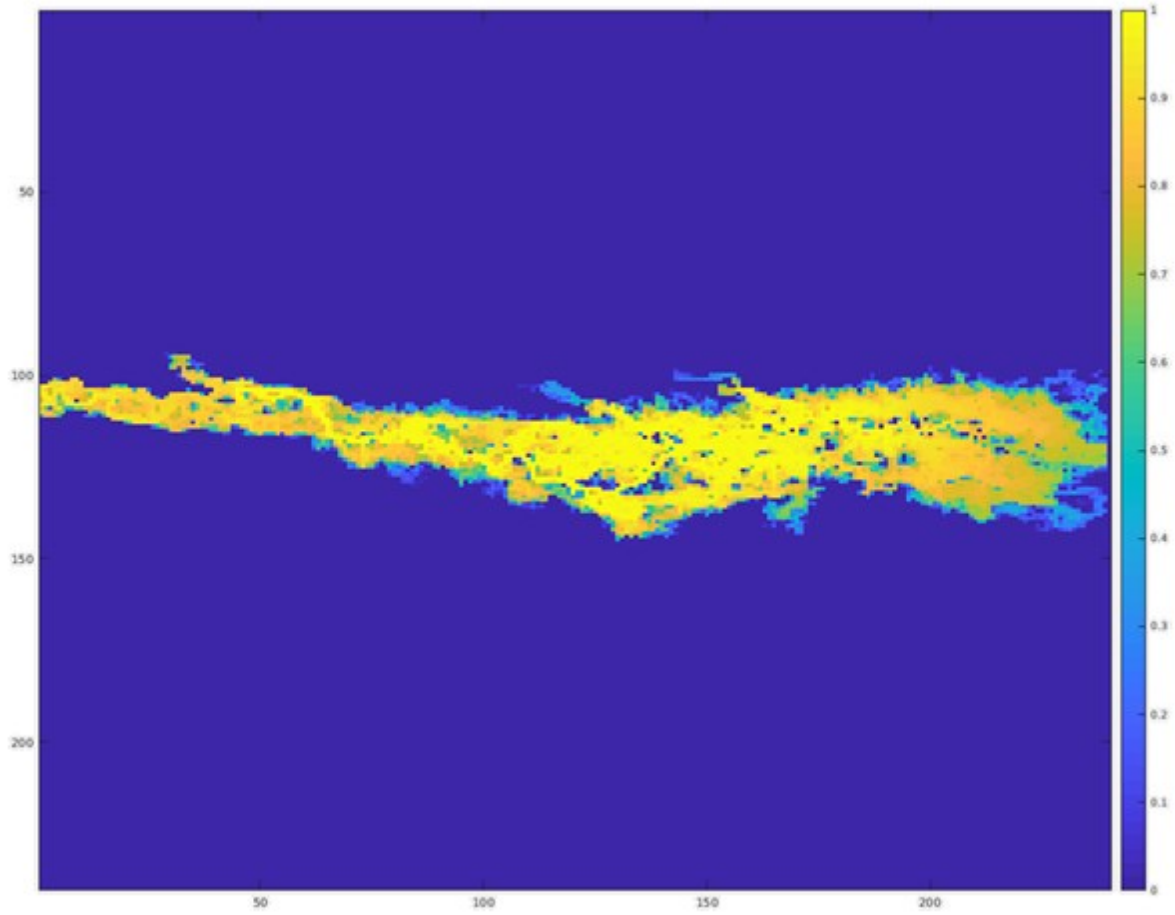


Figure 6. 8 - Heatmap of the Po river basin. Displays the frequency of flooding per modelled cell in the domain. Notable is the lack of flooding in the downstream region across many of the models

The CSI score is limited to a value of 0.7 due to underprediction. This is shown in figure 6.9, where we see a corresponding heatmap of false negative scores. With reference again to figure 6.2, the performance of the model can be compared to the "observed" floodmap available. In this figure, the number of times the model mistakenly predicts no flooding is presented. Clearly there are segments of the Po river which are poorly represented in the model, particularly in the south eastern corner of the modelled domain. There are further groups of false negatives located throughout the main floodplain stem. This is likely due to the uncertainties of the SRTM DEM, as the section of the flood map which is represented by the Po delta has not been included.

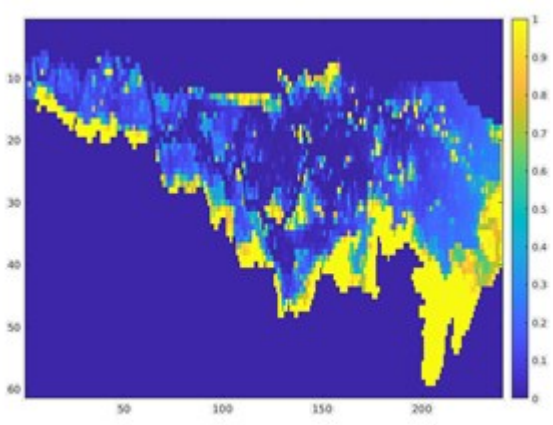


Figure 6. 9 - empirical 2D PDF of false negatives - repeated misrepresentation of the south east section of the basin is occurring.

6.4.1.1. Posterior Distribution Analysis

Having looked at the raw outputs of the results and seeing what some of the primary attributes of the model predictions are, we are now in a position to look at how model performance is linked to parameters. This means taking the MC sampled distributions and calculating the CP metrics. Each individual model run will have a different CP associated with it, which taken together sum to 1. This means we can form CDF's of the parameter's performance and use that as a posterior distribution. The relationship of the CP's and the parameters are presented as CDF's in figure 6.10. The steeper sections of the CDF indicate the areas of greater responsiveness. This corresponds to the peak flow multiplier and reach accumulation threshold most clearly. By contrast, the CDF's of the previous non-informative priors would be a straight line, as their distribution was uniform.

In the figure, the CDF's indicate that most of the parameters are unresponsive, although there is some responsiveness seen in the Q multiplier and the reach accumulation threshold. Except for these, there is little information which can be derived from the CDF's, except that it seems the parameters are equifinal.

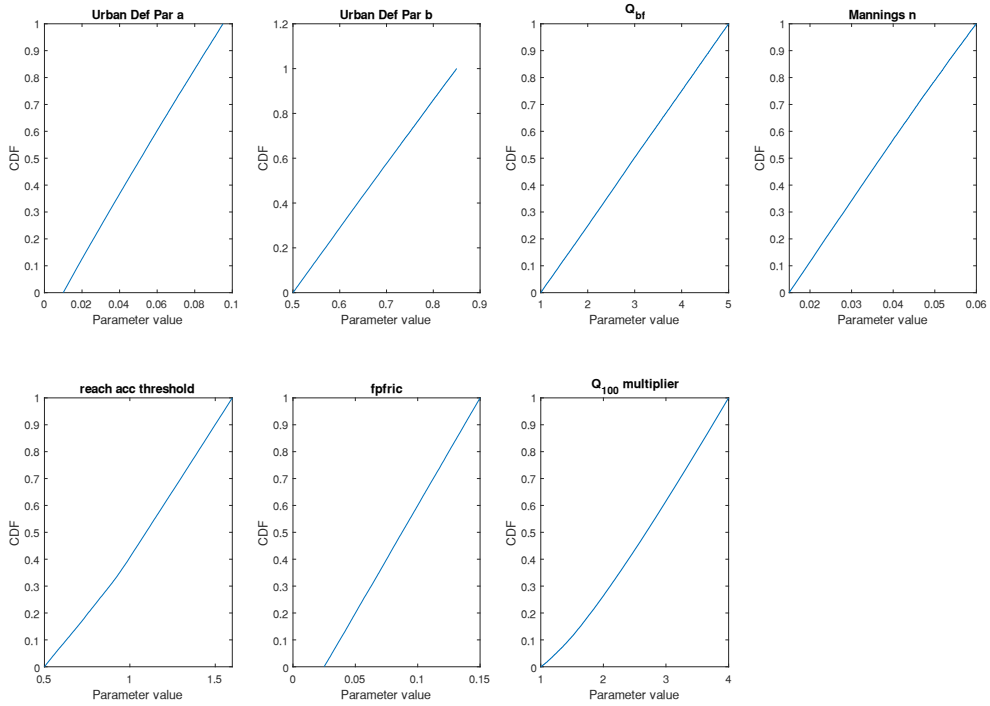


Figure 6. 10 - CDF's of posterior parameter distributions, weighted using the CSI performance metric.

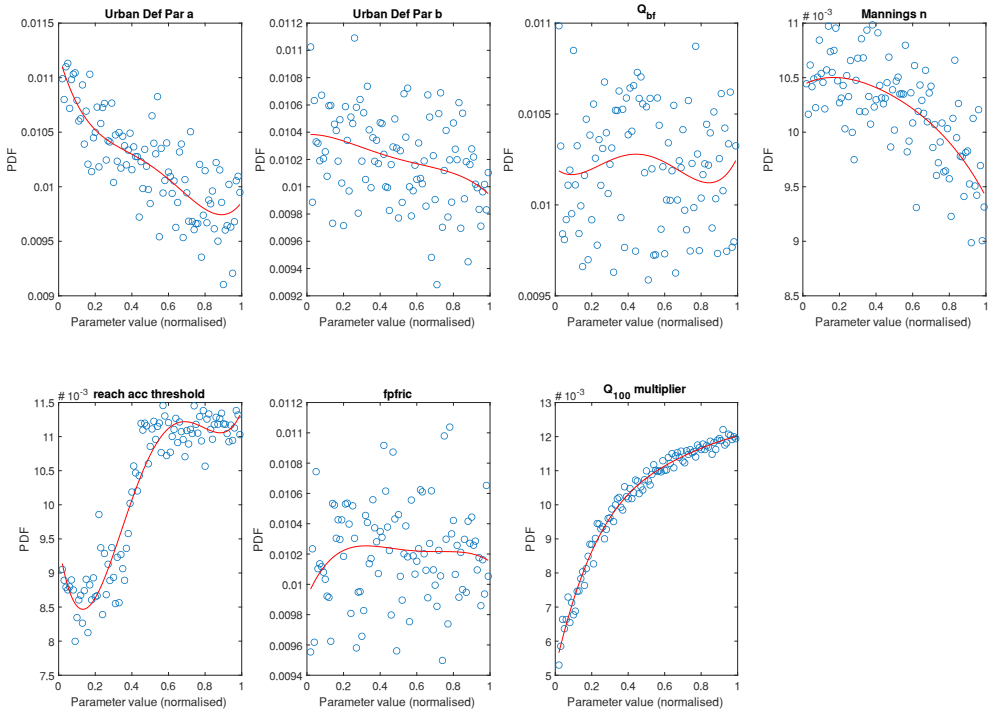


Figure 6. 11 - Posterior PDF's of the parameter distributions, weighted using the CSI performance.

From the CDF's in figure 6.10 we can calculate the PDF's, shown in figure 6.11. As mentioned, these were derived using interpolation and a moving window to counter the noise of a point wise derivation. The interpolation method returns the blue dots, while the moving average window returns the line as a representative of the posterior PDF. Manning's *n* and urban defence parameter show a limited sensitivity, and there are clear effects of the reach accumulation threshold and the Q_{100} multiplier. Notably, model performance continues to rise for the models which have flow values we know to be unrealistically high. This is likely because of the "valley-filling" effect, and the CSI's inability to distinguish past this point. The urban defence parameter *b* and Q_{bf} are likely unidentifiable because they are equifinal. Due to their relevance to the same process of modifying the channel bathymetry, it means that they are all doing the same thing as Manning's *n* and urban defence parameter *a*, likely reducing their importance. Figure 6.11 reaffirms the significance of flow prediction to the model results. As posterior PDFs, the plots show what parameter is likeliest to return a high performing model. Furthermore, the peaks are also the expected value the parameter should take for a skilful model setup. They also confirm the ranking of importance of the parameters. A clear conclusion of the model at this stage is the necessity to account for uncertainty of the regionalised flood frequency analysis, as among the most responsive of the distribution surfaces is the peak flow multiplier. As discussed in section 6.3 however, this was added in retrospect, after seeing the model fail to produce decent model outputs without accounting for the uncertainty of the peak flow.

Figure 6.12 diagnoses the model's predictive ability as a function of how much water is sent into the domain. In fig 12a, it is shown that when we take a top percentage of the best performing models (in the example below it is 30%), their performance is based almost entirely on whether there is enough water in the system. Similarly, in fig 6.12b, it is seen that where the number of true positives diminishes, it is replaced by more false negatives, showing again that the poor model scores are a function of poor inundation extent and specifically underprediction.

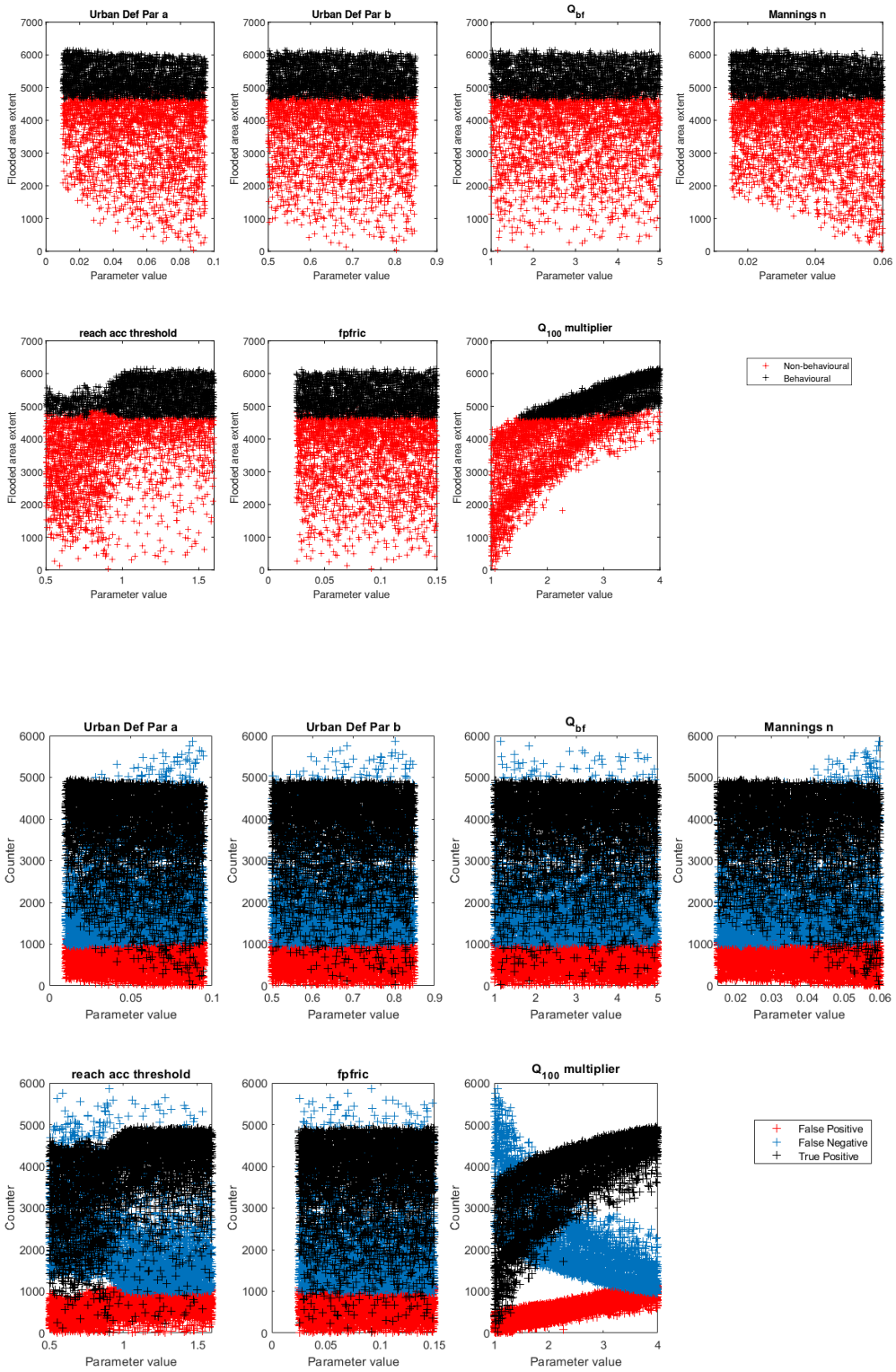


Figure 6. 12 - (a) Dotty plots of parameters recording number of flooded cells per simulation, divided, by top and bottom 50% of parameters in terms of CSI performance. (b) A recording of the counts of true positives, false positives and false negatives present in the modemodel. This confirms the prevalence of underprediction as the source of model failure, and which parameters effect this.

6.4.2. Elevation-based posterior distribution analysis

Figure 6.13 shows the performance of the model at each gauging station, calculated as the error between the observed and predicted elevations at each station. Each gauge has a distinct average elevation error. While the majority of the elevations recorded at Piacenza have very small errors, it also records the largest deviations away from the true elevation. This suggests that the bathymetry is a sensitive issue for the model at this point, and the depths of the river bed as produced by the relevant parameters is responsible for the 15m variation in predicted elevation. At Cremona there is both over-prediction and under-prediction of the elevation. Cremona is situated at the central portion of the basin, which can be seen from figure 6.13 to be the most consistently inundated area of the domain. Borgoforte instead exhibits a systemic underprediction of the elevation, with the average error being 2m below the observed elevations, with a very small variance in this error. Borgoforte is the adjacent gauging station to Cremona. However, this underprediction is potentially due to an error in the locally observed data to estimate the 1 in 100 level at the gauge, which places the peak flow at Borgoforte to be approximately 2,000 cumecs greater than the preceding station, which is a slightly spurious result. Sermide has the largest average error of the 4 stations, as the elevation at this station is continuously underpredicted despite the local estimate of Q_{100} appearing more sensible. This is may be explained by the lack of flood defence representation in the model, as the extensive levee system of the Po does not exist in the DEM used. The effect of this on the routing of water could mean that there is water being held upstream, rather than being routed downstream.

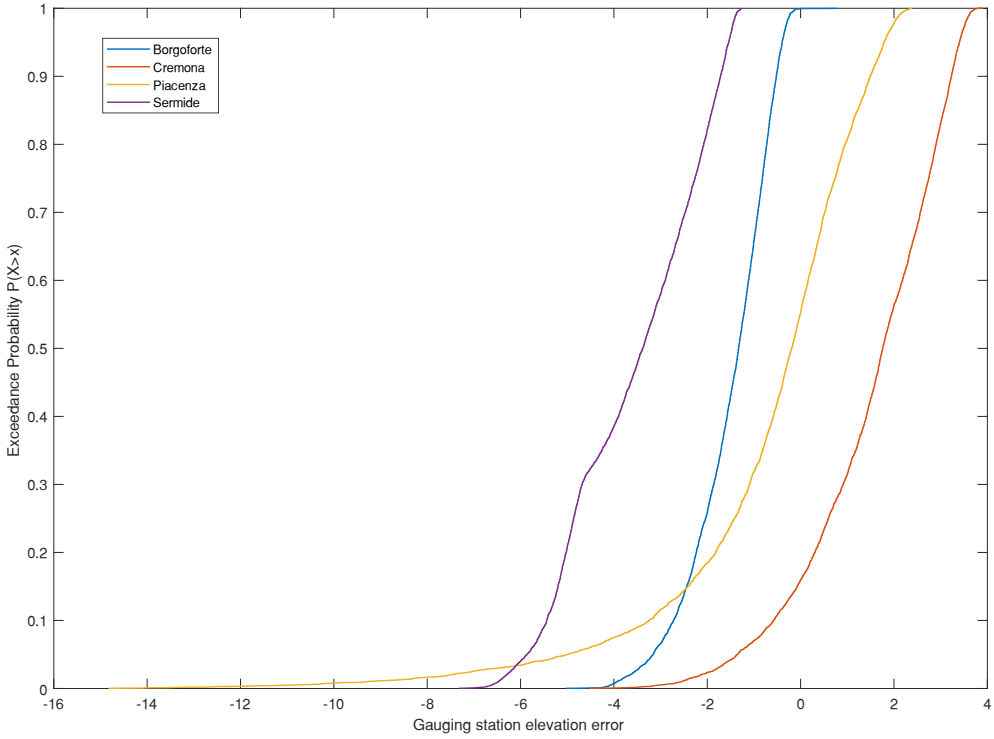
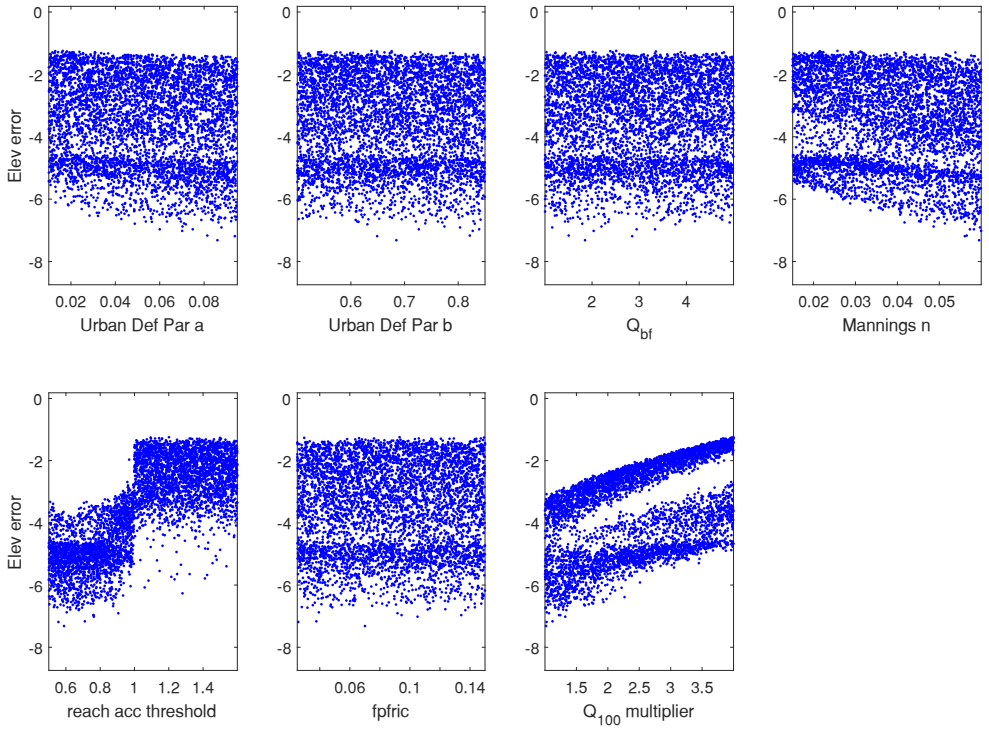
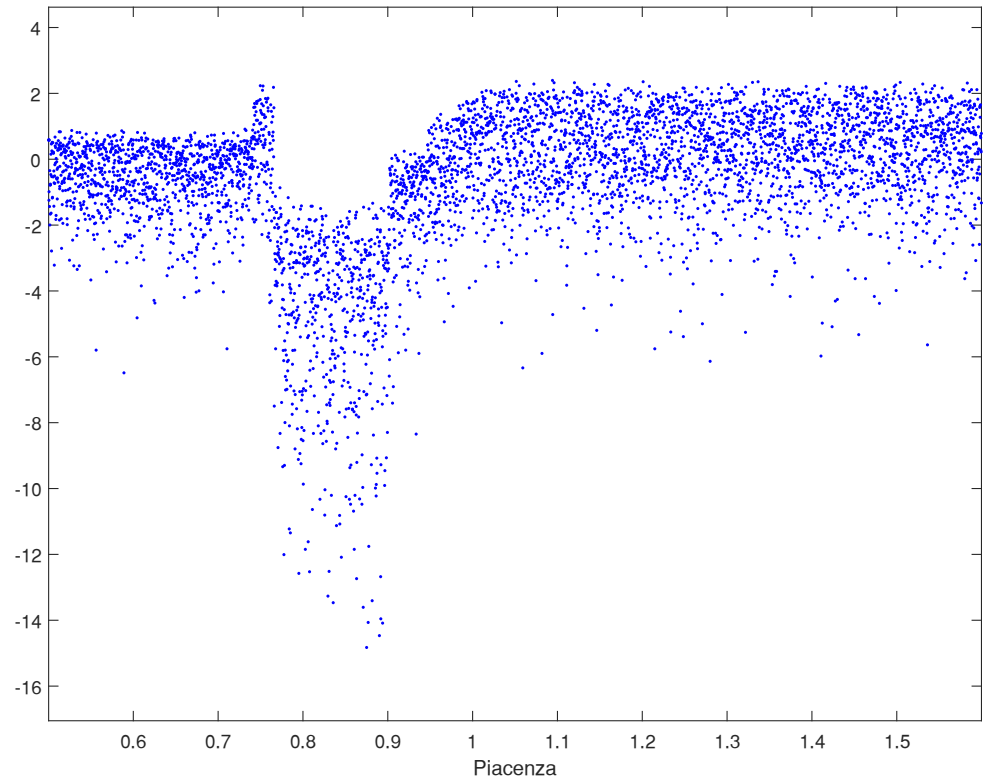


Figure 6. 13 - Elevation errors across the gauging stations for each parameter combination. Elevation at gauging stations predicted for the 100-year event. Piacenza (furthest upstream) shows some of the greatest error but this is prone to greater variability, while Sermide (furthest downstream) exhibits the greatest error with a constant deviation.

The performance of the model at each of the gauging stations can again be compared against the corresponding parameter values. In figure 6.14, this relation is presented using dotty plots. As can be seen, the parameters appear as mostly equifinal. The main exceptions are the Q_{100} multiplier and the reach accumulation threshold, the latter of which produces some highly non-linear effects across the threshold of the parameter at values of 0.8 to 1. Alongside the plots at Sermide, it is apparent that the effects of the reach accumulation threshold are not lumped but are instead highly location dependent. This could also explain the shape of the error CDF shown in figure 6.13. Given that the reach accumulation threshold determines the location of boundary conditions within the model, and the gauging stations are fixed points within the domain, the discontinuous effect on model performance registered in these plots indicates that the model performance is sensitive to the proximity of boundary conditions to the gauging stations.


Elev error
0
-2
-4
-6
-8
0.02 0.04 0.06 0.08
Urban Def Par a
0.6 0.7 0.8
Urban Def Par b
2 3 4
Q_{bf}
0.02 0.03 0.04 0.05
Mannings n
0.6 0.8 1 1.2 1.4
reach acc threshold
0.06 0.1 0.14
fpfric
1.5 2 2.5 3 3.5
Q_{100} multiplier


4
2
0
-2
-4
-6
-8
-10
-12
-14
-16
0.6 0.7 0.8 0.9 1 1.1 1.2 1.3 1.4 1.5
Piacenza

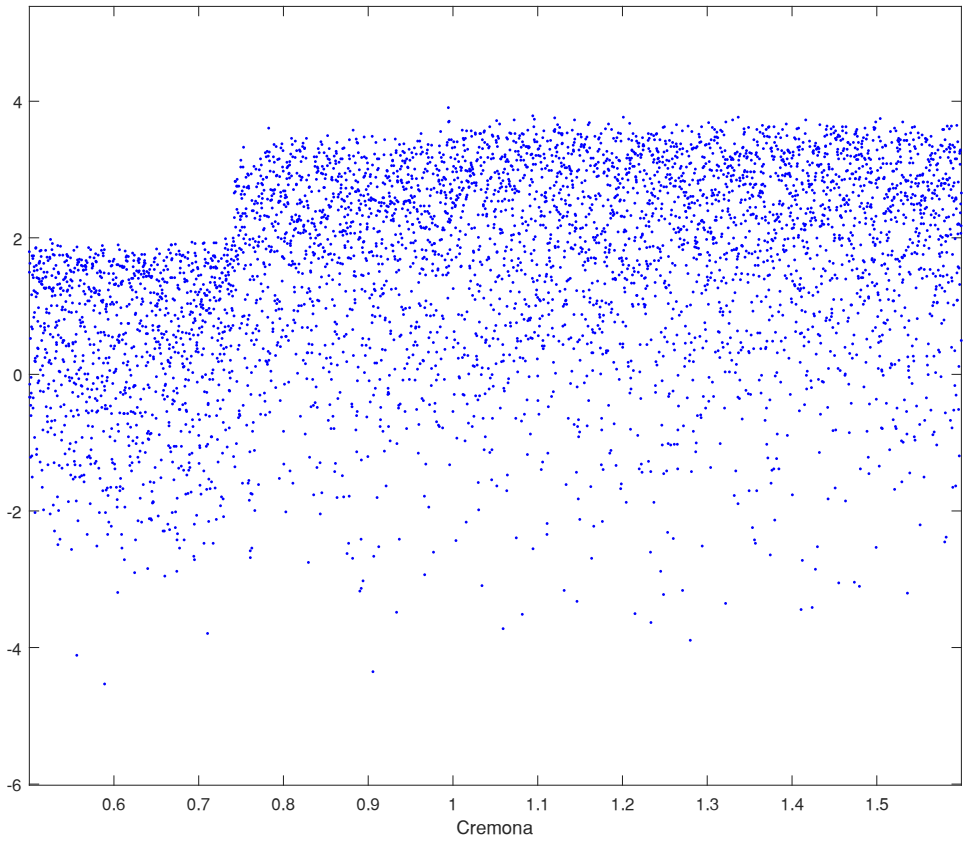


Figure 6. 14 - Tracing the non-linear effects of the reach accumulation threshold: (a) Parameter plots of elevation error at Sermide station. A notable bifurcation in the data is present. (b) Reach acc thresh at Piacenza, (c) Reach acc thresh at Cremona, while there was no response recorded at Borgoforte (omitted).

When we combine the performance of all of the stations and assess the minimum RMSE we get the plot shown in figure 6.15. This shows that the minimum RMSE is slightly under 4m for the entire set of parameter inputs used. The mean error was also plotted alongside the RMSE, to highlight the reason for the systematic RMSE. This shows a prevalence of underprediction happening across all of the models, and confirms the results shown in figure 6.13.

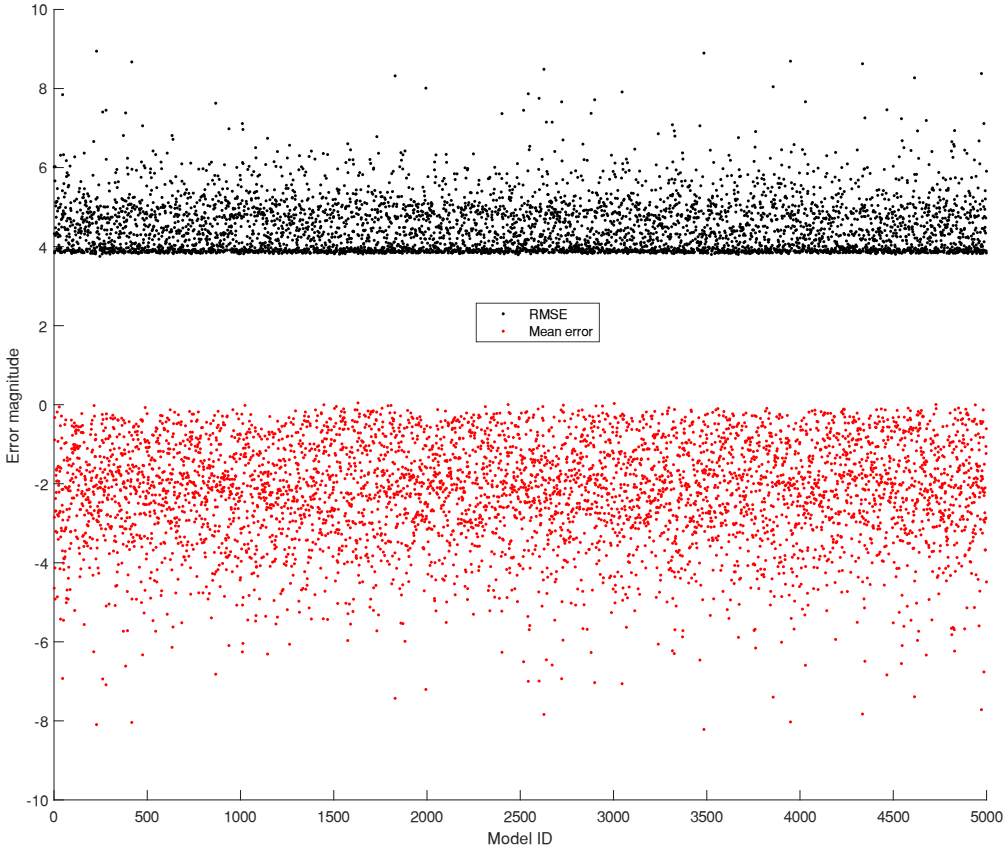


Figure 6. 15 - Lumped elevation errors of the model - comparing mean error and RMSE

6.4.2.1. Posterior distribution analysis

As in figure 6.10 and 6.11, the posterior parameter distributions can be assessed using the CP metric, although this time using the elevation metric instead, shown in figure 6.17. Manning's n shows a more pronounced tendency towards better performance when its value is lower. When looked at together with the floodplain defences, this is an interesting finding, as the lower Manning's n means that water will be routed downstream more quickly and result in lower water surface elevations. However, to do this for a better model performance indicates using a value of Manning's n which is unrealistic. This means there must be another structural reason for this routing, which would be explained by the poor flood defence representation, which is itself a highly important consideration for inundation and particularly the Po river system (Wing et al., 2019).

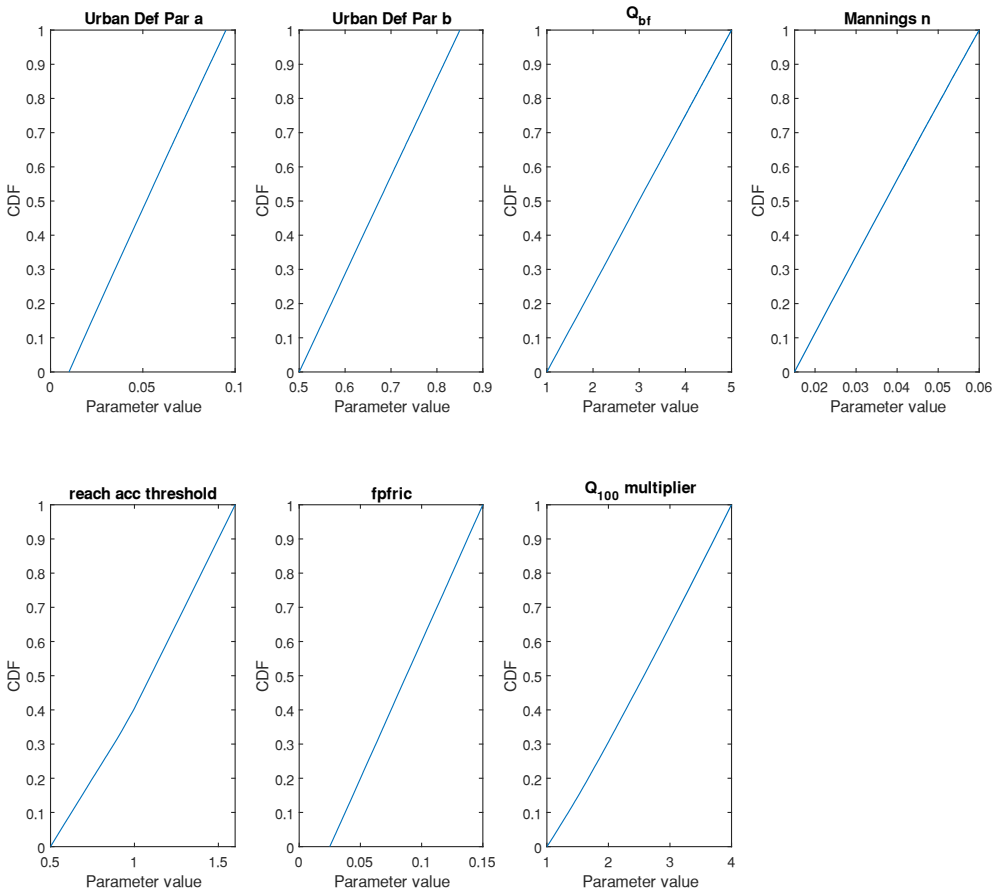


Figure 6. 16 - CDF's of the parameters using the elevation likelihoods.

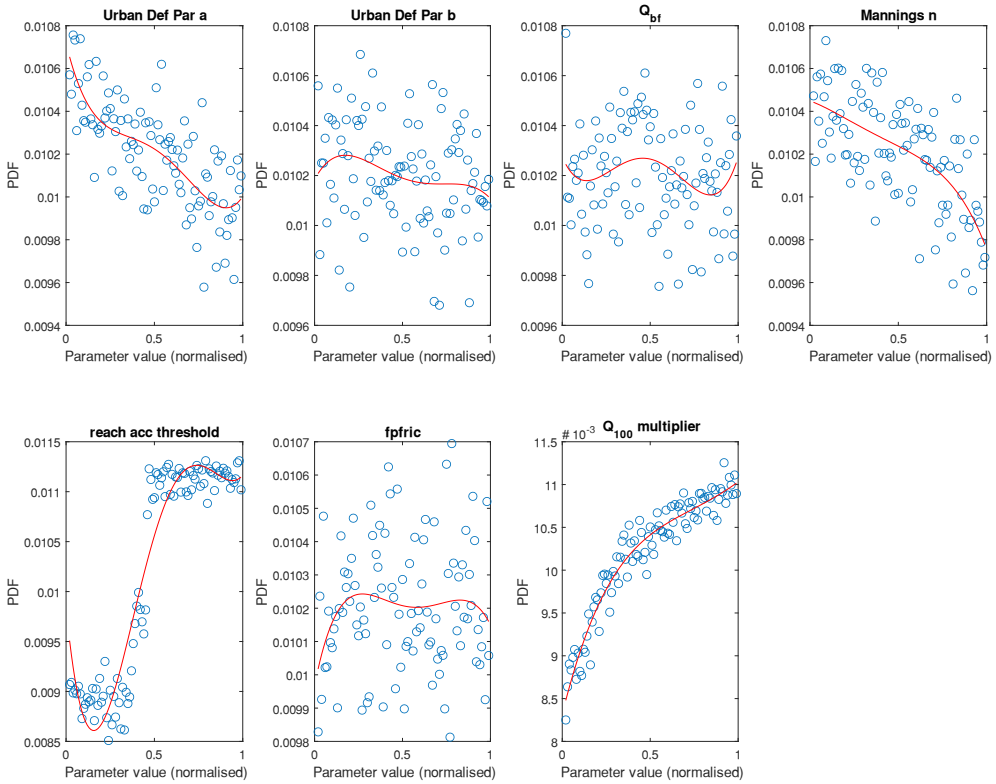


Figure 6. 17 - Posterior parameter distributions using the RMSE of elevation as the weight.

6.4.3. Profile-based CP

Although the primary interest of this chapter is to assess the performance of the models, it is also interesting to note the model's ability to reproduce the state of the Po system. Specifically, the bed elevation along the profile of the channel must be estimated. This ability is in terms of model realism and performance. In figure 6.18, the profile of the Po river at 23 locations is compared with the corresponding cell locations in the model, and those elevations are shown. The model shows a remarkable variability in the modelled bed profile. Below, we see the range of all the different bed profiles produced by the model. The best model approximation of the riverbed is in blue, while the local profile is in red. Note that the blue profile is monotonic, a constraint affecting all the model realisations. The RMSE of this model bed is 1.84m. We can see that the areas of greatest difference at in the steeper regions, such as around stations 5 and 6, as well as at the downstream end when the in-situ bed profile begins to raise up again. Also notable in the model is the great variance of the bed profile due to the inclusion of parametric uncertainty, with some locations having a variation of up to 15m, which is the deviation seen in figure 6.13.

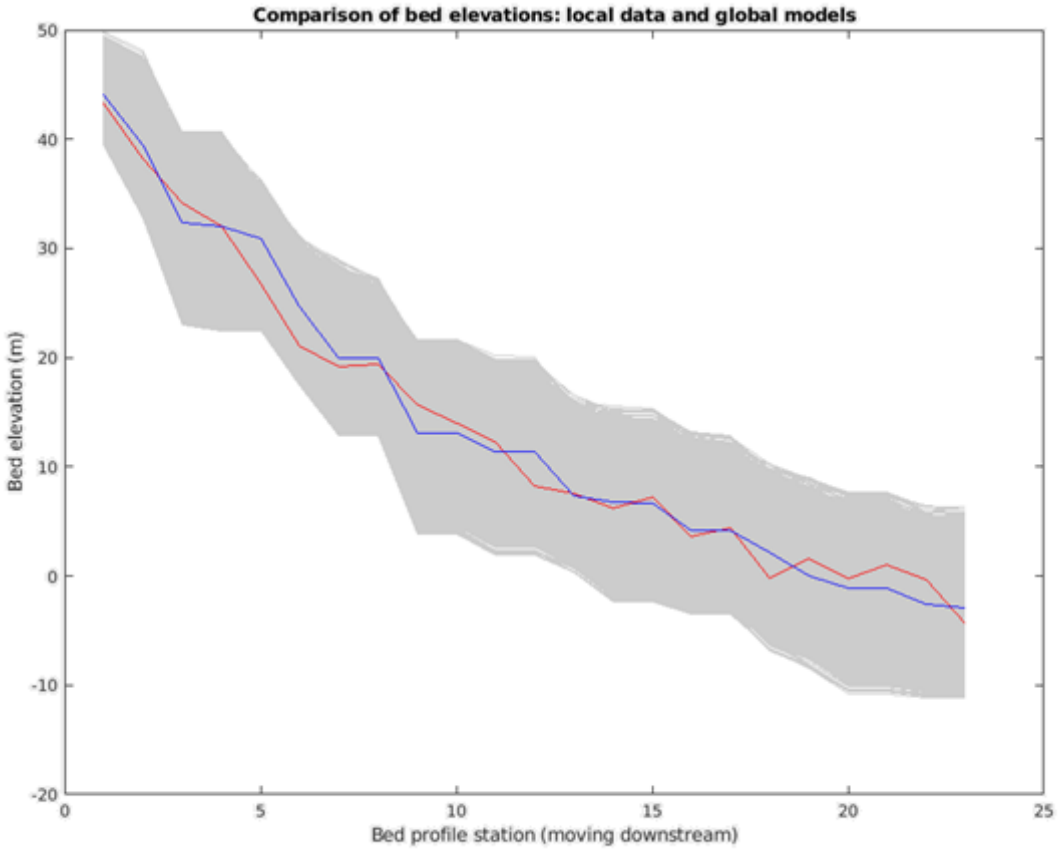


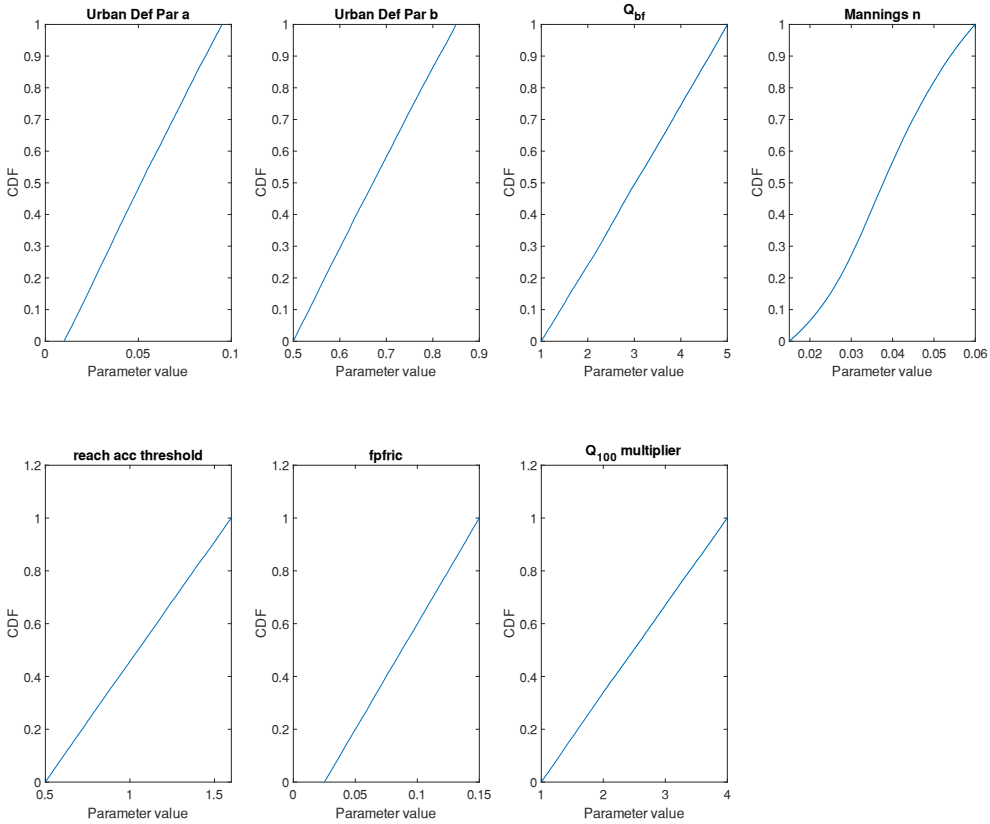
Figure 6. 18 - Approximation of the riverbed profile. In grey shows the range of all modelled beds. In red is the profile of the actual river and in blue is the bed performing river.

The bed profile RMSE does not reduce to less than ~2m, with many models having an RMSE of up to 10m. The inability to reduce the RMSE below this threshold is possibly due to the zigzagging nature of the error structures, which often vary from positive to negative across the locations surveyed. Also likely are that most models fail to capture either the steeper sections of the local profile or the downstream without doing worse on some other sections. Furthermore, as the shape of the channel remains static, rather than changing as in the case of the actual channel of the Po, it is highly unlikely to follow the observed channel bed anyway.

6.4.3.1. Posterior distribution analysis

With respect to figure 6.19, it can be seen that the Q multiplier, floodplain friction and reach accumulation threshold parameters are completely unresponsive. As this likelihood is associated only with the riverbed, this is what we expect to see. We do however see the hypothesis regarding the importance of the defence parameters, Q_{bf} and Manning's n fulfilled. We have seen previously that these parameters have appeared unresponsive and equifinal, but here they are identifiable and responsive. Urban deference parameter a takes precedent over b, because of the extent of the ranges used for the coefficient and exponent respectively. Q_{bf} is less responsive, but certainly Manning's n is the most important.

Interestingly, it confirms our expectation of what the parameter value should be given previous studies on this reach (Domeneghetti et al., 2015). This would indicate that there is a convergence in the model of achieving realism and correctly estimating the bed. The "most likely" Manning's n = 0.03 is in fact the default value used and is often used in many flood studies (Arcement & Schneider, 1989; Yamazaki et al., 2011) Similarly, we would expect this to be a behavioural value for the parameter in the Po flood studies as well, but this would be for the case of modelling the water depth correctly. Therefore, seeing it as the most likely parameter value in the context of the profile elevation is very interesting.


Urban Def Par a
Urban Def Par b
Q_bf
Mannings n
reach acc threshold
fpfric
Q_100 multiplier
CDF
Parameter value

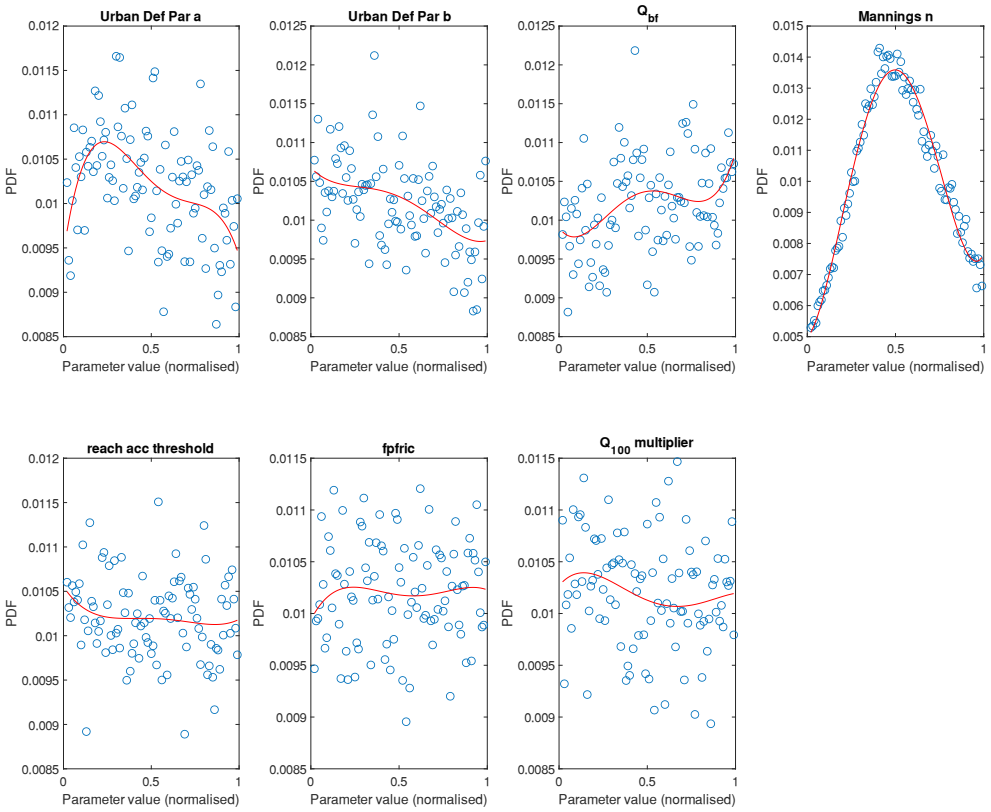


Figure 6. 19 - CDFs and PDFs of posterior parameter distributions for likelihood weighed using bed profile RMSE.

6.4.4. Flow-based CP

Although it seems simple, assessing the accuracy of the Q_{100} in the model is quite difficult. The only way it is possible to make the comparison is using the boundary conditions in the model, which stipulate what the peak flow at that point will be, not how it will propagate downstream. This also has a location which is different to the point of the gauging station. Because of this, it is quite difficult to judge whether the comparison being made is even reasonable, particularly if the closest point where the flow is explicitly known is many miles away. See the previous section in chapter 3 on the setting of the boundary conditions, which explains how this location is parameterised and therefore determined by model processing.

However, by placing a behavioural filter on how close the boundary conditions are to the gauging stations, we can derive some further details out about the model's behaviour. In other words, by filtering the distance of the gauging stations to the boundary condition locations, we can begin to understand the behaviour of some of the parameters better. In

figure 6.19, we see the plot of the parameter “reach acc thresh” against the critical success index. However, each model has been plotted red or blue to represent how close the boundary conditions occur to the gauging stations. The behavioural distance settled on was 0.01 degrees, as this translates to approximately 12km, and at this point a clear demarcation was seen, as shown in figure 6.19. This plot shows that precisely at the threshold of 1, the models can be defined as either behavioural or non-behavioural in this sense.

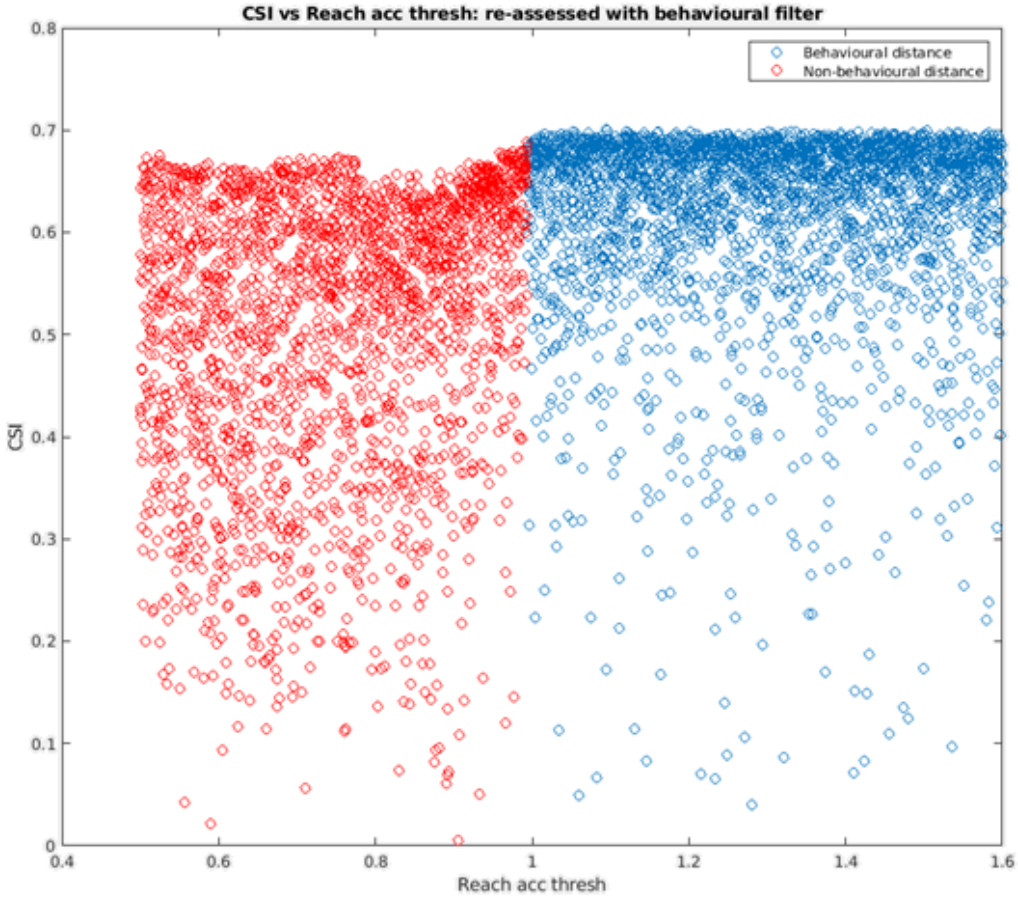


Figure 6. 20 - Parameter plot of CSI vs. reach acc thresh: a filter has been used which deems non behavioural any boundary condition locations which are more than 0.01 degrees away from their respective gauging stations.

Using this distance as a behavioural threshold, the relationship of the Q_{100} error can be compared to the value of the Q_{100} multiplier used. This relationship is shown in figure 6.20. The same behavioural filter can be applied to the estimation of Q_{100}, with the Q_{100} multiplier. The first item to note is that the minimum error occurs within the ranges of the peak flow set,

this being at approximately 1.4. What is also noticeable is the collection of parabolic relations on the plot, corresponding to the parameter and error metric. However, there are multiple parabolas, each one of which will correspond to the different positioning of the boundary conditions. As it has been established that there is a direct relation between the positioning of the boundary conditions and the reach accumulation threshold, then it is clear from the plot below that the two parameters which impact the performance with respect to Q are the peak flow multiplier and the reach accumulation threshold. Hence it is also clear that the best performing model, with respect to the correct Q_{100} multiplier, is first dependent on the boundary condition location being correctly set.

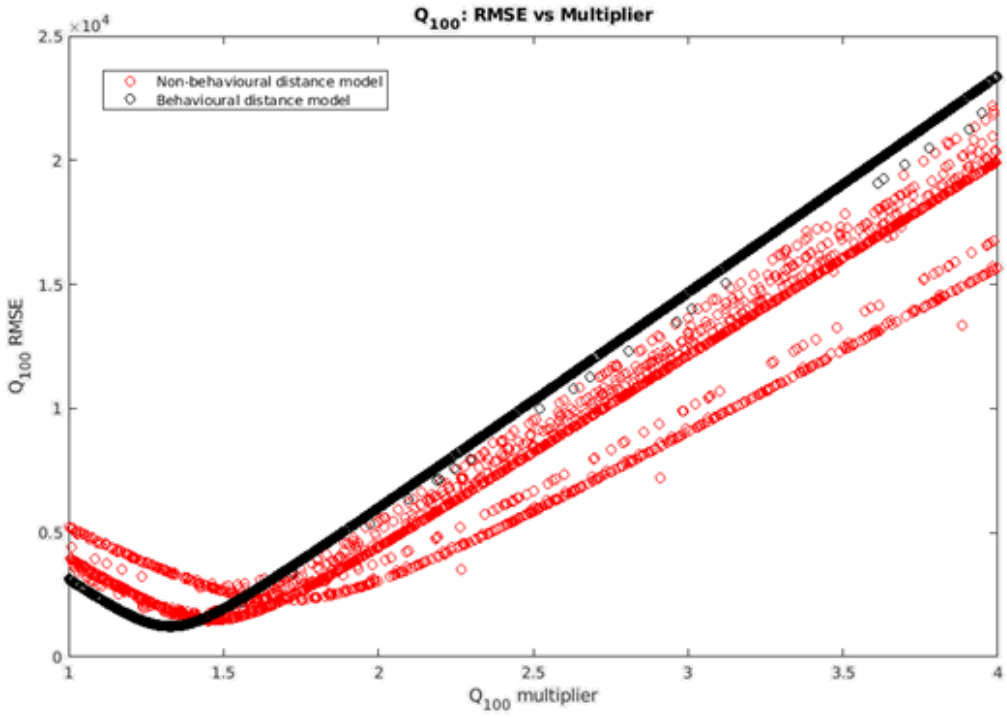
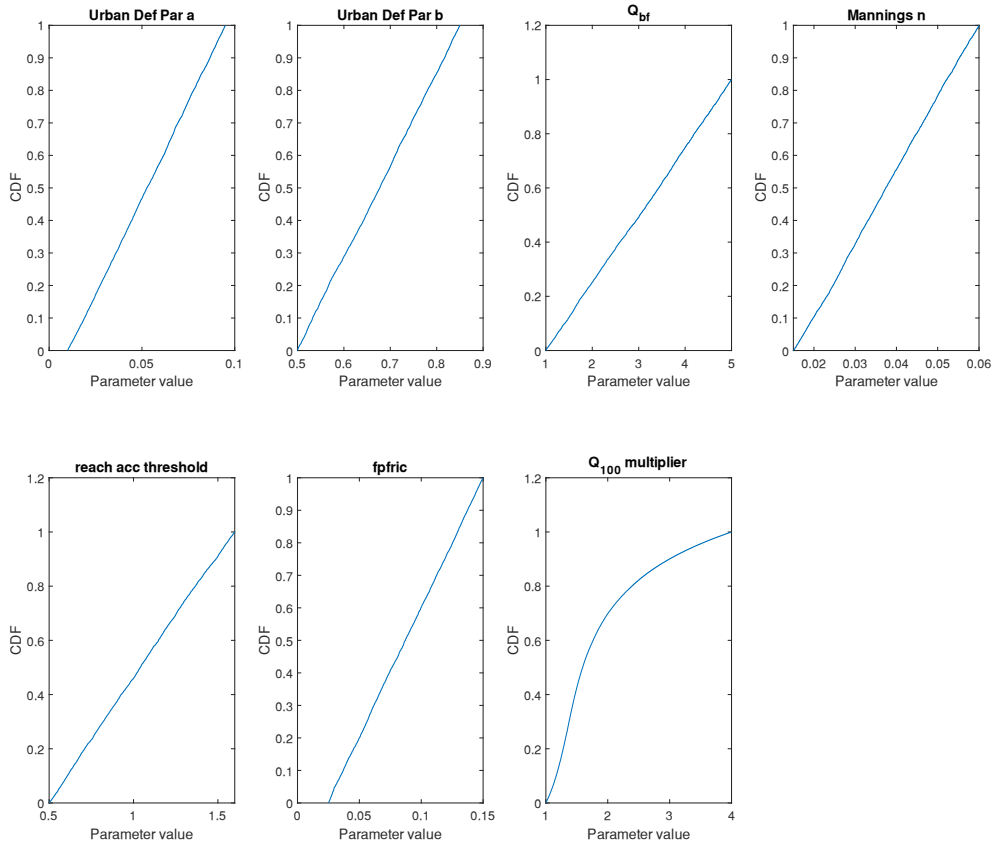


Figure 6. 21 - Plot of Q_{100} RMSE vs Q_{100} multiplier. The values are distinguished based on how close the boundary conditions are to the gauging stations.

6.4.4.1. Posterior distribution analysis

We can also calculate the likelihood CDF and PDF's produced by this performance measure, shown in figure 6.22. Of course, we find there is an obvious response from the Q_{100} multiplier. This alongside previous posterior distributions highlight that the model is more sensitive to underprediction of flow than overprediction. There is also a slight response registered to Manning's *n* in the plots below. Given that we know there is a distance between the gauging stations and the boundary conditions, then the possibility for some small hydrodynamic

effects to exist, in that the distance between the boundary condition and the gauging station, means the effects of friction could change the downstream Q. Furthermore, as the distance of the boundary conditions to the gauging stations is informed by the reach acc thresh, then it is also reasonable to see a responsiveness of that in figure 6.20 as well. In the case of the remaining parameters, given what we know about their effects so far, the peakiness at the extremes of the ranges is likely a spurious result.



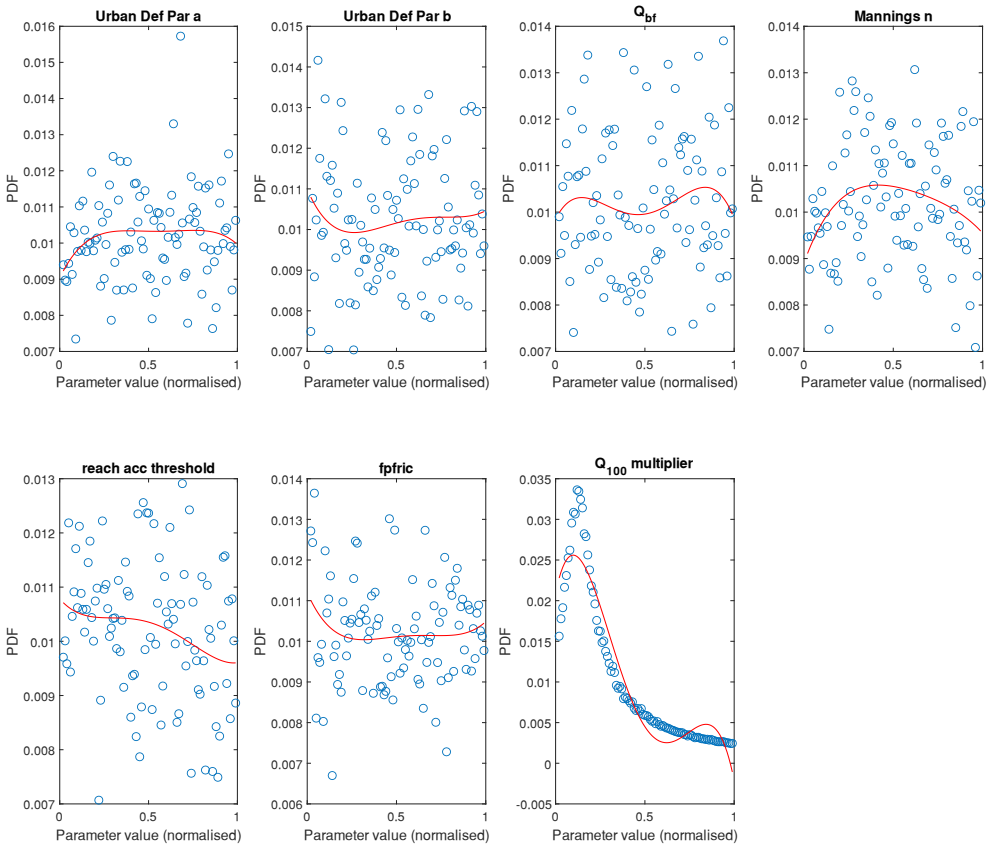


Figure 6. 22 - CDF's and PDF's of posterior parameter distributions using likelihood weighted on Q_{100} RMSE

6.5. Discussion

There are some interesting implications about the behaviour of global flood models based on these results. The next section will consider some of these.

6.5.1. Underprediction of model results

The preliminary test with Monte Carlo sampling of existing parameter distributions all performed very poorly, simply due to the large errors associated with the calculated peak flow. Where we anticipated flows of over 10,000 cumecs, instead we were predicting 6,500 cumecs. This represents an error of about 40%. According to Smith et al., 2015, the median error of the rFFA approach is 56%, thus it is not out of line with expectations given the literature on the extreme Q generation method.

Briefly we will explore potential reasons why the rFFA is underestimating the flows on this river. A strong hypothesis is related to the hydrology of the Po river. Most of the rainfall which

constitutes the peak fluvial flooding in the Po occurs in the mountains, at the edges of the basin (Montanari, 2012). This nuance of local hydrology is processed in the rFFA through average accumulated upstream rainfall. While this means there will be some account of the variability in precipitation from the mountains to the floodplains, it lacks a lot of the nuance associated with the hydrology of the peak flows, themselves a function of the spring time melts rather than simply due to rainfall. This lack of process representation is likely a strong reason for the poor prediction of peak flow.

Once the poor prediction of Q_{100} had been relaxed to allow flows to bracket the observations, our results still indicate that the models are much more sensitive to underprediction than they are to overprediction. In the likelihoods of the area and elevation, the Q_{100} multiplier parameter has no upper limit whereby it reduces model performance by overprediction. This can only be negatively scored when the local flow rate is used as a performance weight and then we see that the parameter is constrained with an upper limit.

6.5.2. Equifinality of parameters and bathymetry

When looking at modelling performance based on the inundation likelihoods, we see a lot of equifinality between the parameters. Here the performance metrics are not discriminatory in the sense that they show the parameters having any significant impact on model behaviour, as our screening would imply. This is likely to do with the fact that the parameters of urban defence (a and b) and Q_{bf} essentially perform the same function; that is, to modify the riverbed through the inversion of the Manning's equation. This is the reason we chose to look at the model's ability to recreate bed elevation and were reassured with the results which show these parameters as identifiable in this regard. Overall, what these results tell us is that calibration of the model ultimately hinges on the bed elevation and calculation of Q_{100}, while the friction coefficient also needs decent estimation.

6.5.3. Role of the reach accumulation threshold

The reach accumulation threshold has unexpectedly emerged as both highly sensitive and fascinating on the influence it has on the modelling domain. As a threshold, it determines what constitutes a reach within the domain. Each one of these reaches is individually modelled and then aggregated together to create the hazard maps. As LISFLOOD-FP is hydrodynamic and mass conservative, it does not add extra water along the reach in the way it's setup here (Bates & De Roo, 2000; Bates et al., 2010); instead the model discretely adds

water at each new boundary condition, defined per reach. As this is based on the incremental increase of upstream accumulated area, we can see how this could have a remarkable effect upon the model.

The discrete effect of the parameter, which has a qualitative shift in its effect around the value of 1 has been puzzling. However, on reflection, we are able to reach an understanding for this apparently mysterious behaviour. As we know, the reach accumulation threshold determines the ratio of US and DS accumulation between a reach within a river system. In the case of the Po downstream basin, we are working at the scale of 10^4 km^2 accumulation. This means that the accompanying increase in accumulation over a maximum of 200 km is likely to be something like less than 1% deviation in the value. This would mean that any value less than 0.99 would revert to the maximum reach length for calculation of the boundary conditions. This means that there is a fundamental issue of scalability with the reach accumulation threshold. While the default assumption of 20% change in accumulation is no doubt suitable for a sizeable proportion of regional reaches, it will not be suitable for both smaller and larger rivers. There is an implicit assumption in the parameterisation that the relationship of reach length to accumulation is linear, which is not actually the case and has led to the unexpected behaviour which is seen here.

6.5.4. Incorporation of local data

It would be beneficial to utilise local data with the GFM to produce effective results. Although it can be claimed to have a global scope, we are still not able to say that global estimates can be made effectively without being conditioned by more local scale data. This was also true for the closely related fathom US model, which was reliant on local data pertaining to the flood frequency analysis, to make it more accurate than it at the global scale. Hence, we are left with the question of how transferable much of the model skill is to ungauged and data-sparse regions, and whether we can find other sources of information or data to help condition model output (Fleischmann *et al.*, 2019).

It is also very clear from this study there is a need to better incorporate observational uncertainty. Despite using multiple sources of data, we have been unable to properly quantify a “degree of belief” to associate with our models (Nearing et al., 2016), despite having better understanding of the posterior parameter distributions. This degree of belief could not be reached because there was not enough information about the provenance of the data which

was used. For instance, the ratings were products of higher resolution averaging, meaning that the uncertainty could not be parsed in any meaningful way. Despite the shortcomings of the attempts at quantifying the uncertainty here, it does not detract from the fact that all other GFM's should look to incorporate observational uncertainty in the same way, or risk underestimating the total uncertainty.

6.5.5. Performance and realism

A broader topic of interest related to modelling is whether the higher performing models also have realistic parameter values. This in turn implies that the physical processes are correctly understood and the model is able to represent all of the mechanisms correctly. We further tested the model in this sense by assessing the ability to recreate the observed river channel bathymetry. Given the assumptions made regarding the bathymetry in the global model, it wasn't expected to show such a strong link between realism and performance, but the response of the Manning's *n* is quite clearly in line with literature expectations, and we also see some identification of the other parameters related to the bathymetry. This coherence of realism to model performance is reassuring for future analyses that this GFM continues to be grounded on reasonable assumptions and provides decent representation of the dominant hydraulic processes. However, given the present equifinality and the unquantified effects of input uncertainty, there is more work to do on the interactions of uncertainty between different parts of modelling chain, to better identify how the uncertainties are propagated to the final output (Wagener et al., 2003; Wagener & Gupta, 2005), which would involve repetition of this experimental design at other locations and with a greater variety of local data streams.

This consideration is also important with respect to the limitations of the benchmarking data as locally modelled hazard, rather than the preferred "ground truth" dataset. While CSI values of 0.7 is satisfactory, this has to be caveated with the fact that there are likely errors between the local hazard data and the actual ground truth 100 year event (if we assume that such an event could occur). In the local to global comparison, we see that the global model underpredicts, particularly in the downstream area. However, it could be that there are a different set of assumptions underpinning the local estimate of hazard at this location, such as the presence of coastal effects or another tributary. Alternatively, we could suggest that there is some degree of overprediction in the local model compared to the global model, as

the global model is suggesting a hazard layer for a basin wide 100 year flood event, while the local hazard layer could be an amalgamation of many smaller 100 year events all brought together, which are even less likely to happen at the same time. While this is a possibilities, we have supplementary information from the other performance metrics that there is some systemic underprediction due to the prediction of peak flows in the rFFA. This could lead us to conclude that the actual flood risk in the Po basin is somewhere between the local and global model.

6.5.6. Po defences

Despite there being no clear representation of defences within the model, the flooding was all contained within what is the actual limits of the 200-year flood levees. Despite this being a promising result, the effects of the levees on downstream routing were found to be lacking. However, this was in the case of the Po river, which has a very prominent flood defence scheme not likely to be matched in many ungauged basins. However, this still represents an area of work for further study, and given the recent global dataset on worldwide defences, this is something which can proceed (Scussolini et al., 2016), and other studies assessing the flood defences of large scale rivers like the Po presents the possibility of this soon being covered in detail at ungauged basins (Wing et al., 2019).

6.6. Conclusion

This paper presents a method for assessing performance of the Bristol GFM, accounting for the uncertainty of the parameters and the local data. Using a variety of performance measures, the model was tested to see how well it was able to represent the various sources of local data, and what local data were needed to make that possible. As a proof of concept, global models do have the capacity to be locally relevant, but this is predicated on the availability of local information. A key source of information not obtained here but which would be highly important in future studies is the representation of floodplain defences.

It was found that the main source of uncertainty in this model is the calculation of the peak flow, which here was the decisive factor of whether the model was able to perform well or not. By accounting for the error of the flow and introducing a parameter to represent that uncertainty, we found the tendency of the model to be biased against observations and benchmark modelling results, and that over-prediction was not so clearly detected. Another

key process of the model is the reach accumulation threshold, which here has been shown to have some highly non-linear effects. This is an automated process in the model which is responsible for defining upstream boundaries and number of reaches. The importance of this parameter will be linked to the size of the river being simulated. It has also been found that how a GFM decomposes its domain for the purposes of computation has direct and non-linear effects on the model output.

Incorporating riverbank elevation and discharge of the flow as further performance metrics meant we were able to constrain our best performing models with a sense of realism, and it was found that sensible parameter values were chosen in the posterior parameter distributions. The finer the resolution/representation, the more this will need to be done.

While this method was tailored specifically for the Bristol flood model at the Po river, the approach taken could be easily adapted to other flood models at other locations and presents a contribution to the framework of global flood model validation. From such studies, the inter-related behaviour of the parameters can become known and help to clarify what the differences in the model behaviour are.

6.7. Post-script

In this study, we have discovered some interesting features of the Bristol GFM. While some information has a general value, it is still the case that this work could benefit from being extended to other basins, as undoubtedly the uniqueness of place of the Po basin presents a very specific picture of the model, which still needs to be generalised. Regardless, enough information about the behaviour of the model, and its most poorly performing section (The Q100 estimation and uncertainties with this) could be identified and will be examined in more detail during final results chapter of the thesis.

This concludes the uncertainty chapter of the thesis, which follows on from the specifications made in the preceding “sensitivity” chapter. Furthermore, it also concludes the Monte Carlo analyses of the model as a whole. From these we have identified a portion of the model which requires the most obvious attention, which we will now address in the “refinement” chapter. This refinement will centre around the peak flow estimation component of the model. While

there will still be quantification of uncertainties, it will all be based on a specific component of the Bristol GFM – the regionalised flood frequency analysis.

7. Research Chapter 3 - Assessing uncertainty of model peak flow predictions, and improvement by using more appropriate climate indices

7.1. Introduction

This thesis has so far been interested in the practice of global flood modelling as seen through the sensitivities and uncertainties of a single model, with the additional aim of finding parameter behaviours which are applicable to all global flood modelling practices. The sensitive parameters and processes were found to relate to the method of breaking down the domain; bathymetric and defence representation; and channel roughness. Despite these general conclusions, the previous chapter also found that the representation of uncertainty in the peak discharges is lacking, as wide parameter ranges were unable to capture the floodplain extent in a test case for the Po river. We therefore move to looking at a specific aspect of the Bristol GFM which we deem to be essential, the regionalised flood frequency analysis (rFFA) method used to estimate peak discharges of a given return period. This reduces the scope of the previous chapters to a single component of the modelling chain. However, it has been demonstrated that of the entire model it is this aspect which most urgently requires improvement.

This chapter will develop the rFFA approached used in the Bristol GFM, which is one core aspect of the whole modelling chain and was seen as crucial in Chapter 6. Firstly, we shall quantify the uncertainty of the method in its current application and then examine how the method can be improved. Following this, a secondary approach will be proposed and evaluated, and the relative abilities of each method will be assessed, in terms of their predictive ability and the definition of uncertainty.

This chapter quantifies uncertainty within the rFFA component of the Bristol GFM, by considering the Global Runoff Data Centre gauge dataset from which the method is derived. Using a split-sample method, a training data set was used to perform the various fits and regressions that make up the method. The test data sample was then used to evaluate the performance of the method, and the levels of uncertainty. Following on from this we considered alterations to the methodology in order to constrain the uncertainty generated, and deliver improved estimates of peak flows for a given return period. This was achieved by exploring some of the methodological faults in the current method, such as improving the schemes of regression and clustering as well as incorporating a different climate classification system. This change of the climatology was based on the assumption that the Koppen Geiger

zones might be too coarse a representation of climate for it to be an adequate predictor of peak discharge, while a more recent classification scheme represents a possible solution.

7.2. Part 1: Estimation of peak flow uncertainty in old method

There are two parts to the final chapter. In this first part of the chapter, there is an examination of the current regional flood frequency analysis as described by (Smith et al., 2015), which assesses the skill of the method and how well the uncertainty can be quantified given the current structure. Following this, a second section considers and evaluates a number of changes to the rFFA with the aim of improving discharge predictions and the predictive uncertainty.

7.2.1. Introduction

Regionalised flood frequency analysis is currently not a widely used method for many amongst GFMs. Instead, GFMs often rely on hydrological models and propagating rainfall estimates through routing schemes (Winsemius et al., 2013; Yamazaki et al., 2011). For the purposes of risk estimation, flood frequency estimation is the common choice, although the use of gauging data and regionalising based on catchment characteristics is more commonly done at the reach-scale, as the flood frequencies are instead derived from the outputs of hydrological models or rainfall extremes at the global scale as discussed in the literature review. Despite its prominence more at the reach scale, regionalised flood frequency analysis is viable, as once undertaken it can provide predictions for any basin on the planet, assuming an adequate sample is used. It is yet to be seen whether the stochastic approach of rFFA will succeed the process-representation of the hydrological modelling done in competing GFMs, but certainly the uses of rFFA already done raise questions about the efficacy of the method, and to what degree one can be certain of the predictions that it makes.

There are a multitude of FFA approaches available to the modeller, which have varying reliability and estimates of uncertainty, given their application and data availability (Renard et al., 2013). For example, continuous simulation of a hydrological model has been used to assess the uncertainty of flood frequency estimates at the local scale (Cameron et al., 2000a; Cameron et al., 2000b; Cameron et al., 1999). However, conducting an uncertainty analysis of an rFFA for across the globe presents a different challenge, in that the appropriate mapping of physical characteristics onto predictor variables can vary with location. Given the variability

of the global hydroclimatic environment, this adds further complication to an already significant level of uncertainty. By contrast, the rFFA only utilises 3 variables, namely climate, area and average rainfall. As such, it will not be possible to capture the full range of hydroclimatic variability with such a limited description, as other drivers of flooding and high flows are not represented, such as seasonality and storm intensity (Carey et al., 2010; Zaman et al., 2012). This is not to suggest that the concept of the approach is flawed, as regionalisation is a good way to overcome the problem of uniqueness of place and data scarcity in a GFM, but there is already the caveat that the mapping of physical behaviour using only 3 variables is likely to have limited skill in prediction of extreme flow. For example, a methodological source of uncertainty in the model will be the use of these variables for regression. This assumes that some aspects of basin flow, such as the mean annual flood (MAF) can be reducible to only the area or precipitation. As is the case with almost all regressions, this will not be a perfect correlation, and the distribution of error predictions within this regression is a reflection of the uncertainties present in the method.

What follows is a description of the current method applied in the Bristol GFM, and the decomposition of the method into its various constituents. It is each of these components which can be analysed in a Monte Carlo framework using step wise sampling, which means that the resultant uncertainty of the entire method can be quantified.

7.2.2. Methodology

7.2.2.1. Flood frequency analysis

The method used in the Bristol GFM and applied throughout the applications of this model to date has been based on a regionalised flood frequency analysis. The classical approach of flood frequency analysis for a gauged basin is the use of annual maxima of observed flows (here called AMAX), or some other peaks over threshold value (Robson et al., 1999). Assuming the sampled series of maxima is an adequate representation of the site in question, the exceedance probabilities of these maxima are computed. Following this, the magnitude of the event size is fitted using an extreme value distribution, such as the Gumbel distribution. It is a probabilistic method in that a distribution is fitted from the peak flows to their probabilities, allowing the extraction of return periods for a given peak flow (NERC, 1975). In this way, it is critical for hazard mapping, which produces layers for a given return period, or flood frequency. As one turns to look at the application of these methods to ungauged basins,

the regionalisation component emerges. Without a time-series at the location of interest, the only solution is to take a surrogate catchment and transfer information such as catchment descriptors to the ungauged basin, relying on the surrogate catchment's time series and making a strong assumption about the similarity of both. This is often referred to as "trading space for time" (Singh et al., 2011).

Once the AMAX series is isolated, one selects a type of distribution which most easily able to represent the data. In most cases these distributions are non-normal and focus on the tail component for the extreme flows. This means that distributions such as Gumbel's and Weibull's are often used. In the approach of Smith et al., 2015 the distribution fit is done with a generalised extreme value (GEV) distribution. This means that each local site's fit to the time series has an optimal parameter set. While this means that the fitted parameters are easily transferable to other sites if appropriate, it also means that the method is totally data-driven, and therefore our information sources have to be reliable.

7.2.2.2. Index-flood method

In applying the technique of flood frequency analysis across the global scale, a key problem we face is that of catchment scalability. While we may have two catchments which have a similar distribution of return period responses, the actual magnitude of the extreme flow will be different because of the relative sizes of the catchments and as the model works in the flow units of m^3/s instead of mm/hr, we encounter this issue. Instead, scalability is addressed with the index flood method (Kjeldsen & Jones, 2007). In this, return period flows are calculated from mean annual flood (Q_{MAF}) and a growth factor (GF). Hence Q_{100} would be calculated in the following way:

$$Q_{100} = Q_{MAF} \times GF_{100}$$

Where the components of Q_{MAF} and GF_{100} are calculated by:

$$Q_{MAF} = \frac{\sum_{n=1}^{N} Q^{AMAX}{}_n}{N}$$

$$GF_{100} = GEV^{-1}(0.99, \mu, \sigma, k)$$

Where Q_{MAF} is the mean of the AMAX values over the entire time series, and GF is the growth factor, which scales the mean annual flood to the return period. GEV^{-1} is the inverse generalised extreme value distribution function, which returns the growth factor, given the

parameters of the function, and our desired cumulative exceedance probability. Hence for GF_{100}, this probability is 0.99.

The growth factor is functionally the same shape as what is used in other traditional flood frequency methods, except that it is normalised to the mean annual flood. This means that the shape of catchment response can be retained, but the scale is fitted to something more appropriate. Hence this method is suitable for applications in ungauged basins, where response and magnitude are separate variables which can be estimated from varying sources of information. What these sources are and how they are applied is covered in the next section.

7.2.2.3. Regionalisation

To produce flood maps at the global scale, there are many thousands of river reaches that require peak flow estimation for a given return period. Even in densely gauged networks, a significant proportion of reaches will not be gauged and thus have no streamflow information. The majority of the world's catchments are currently ungauged and even if a substantial observation programme was launched it would take many decades to observe a long enough time series for a standard flood frequency analysis to be undertaken. This means to derive information about flood frequencies, we are dependent on the information we have at the much smaller subset of gauged stations, and extrapolating the information effectively, via regionalisation. For the Bristol GFM, this means estimating the MAF and growth curves from an available global dataset that covers different geo-climatic zones.

The estimation of MAF and the growth curves in ungauged basins are all derived from the estimates of gauged stations in the global runoff data centre's (GRDC, 2018) database. Along with the AMAX time series, we also have information about the catchments, which we can compare with our ungauged basins to perform the regionalisation. For each of the ~5,500 catchments within the GRDC database, we have the location, average annual rainfall, upstream area and Koppen-Geiger climate type. The Koppen-Geiger classification scheme is a climate typology which splits up the globe into zones and are then considered to have homogeneous climates – as a result of the value of certain climatic characteristics such as temperature. The technique of Smith et al. focusses on using the 3 variables of climate, area and rainfall to perform the regionalisation. Another important caveat to add is by using only these characteristics, there is no mechanism to understand whether the flows of the GRDC at

naturalised, which has a significant effect on the estimation of extreme flows. Given that this information isn't known, there is the strong assumption of all flows being naturalised.

The first step of the rFFA is to assign each catchment into its corresponding Koppen Geiger climate class. This means that the fitting of the MAF and growth curves all occurs with respect to the Koppen Geiger climate zones. Once each climate zone is distinguished, an individual MAF surface is derived for each, using area and rainfall as predictor variables. The surface is derived using a 2-step procedure. Firstly, a power relation is assumed between area and Q_{MAF}:

$$Q_{MAF} = a.AREA^{b}$$

Where *a* and *b* represent coefficients, which are estimated using a least squares optimiser. This fit is shown in figure 7.1, for 50 individual gauges.

However, alongside this there is an assumed relationship between MAF and precipitation, which aims to maintain an increasing MAF as precipitation gets larger. This can be written as:

$$\log(Q_{MAF}) = c \times RAIN + d$$

Where *c* and *d* are coefficients in a first-degree polynomial which links the log of mean annual flow to rainfall, a fit which is demonstrated in figure 7.2. How these two fits are incorporated into a cohesive surface will be covered in the next section.

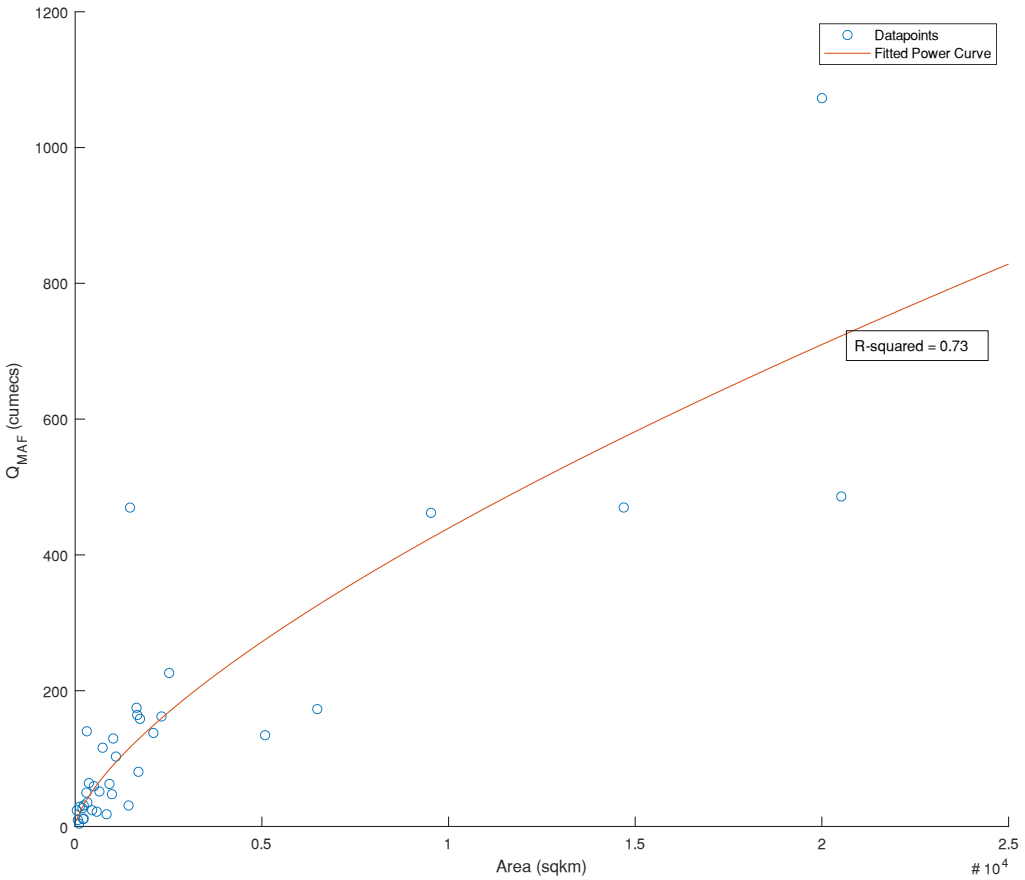


Figure 7. 1 - Example of a power curve used to correlate MAF with area.

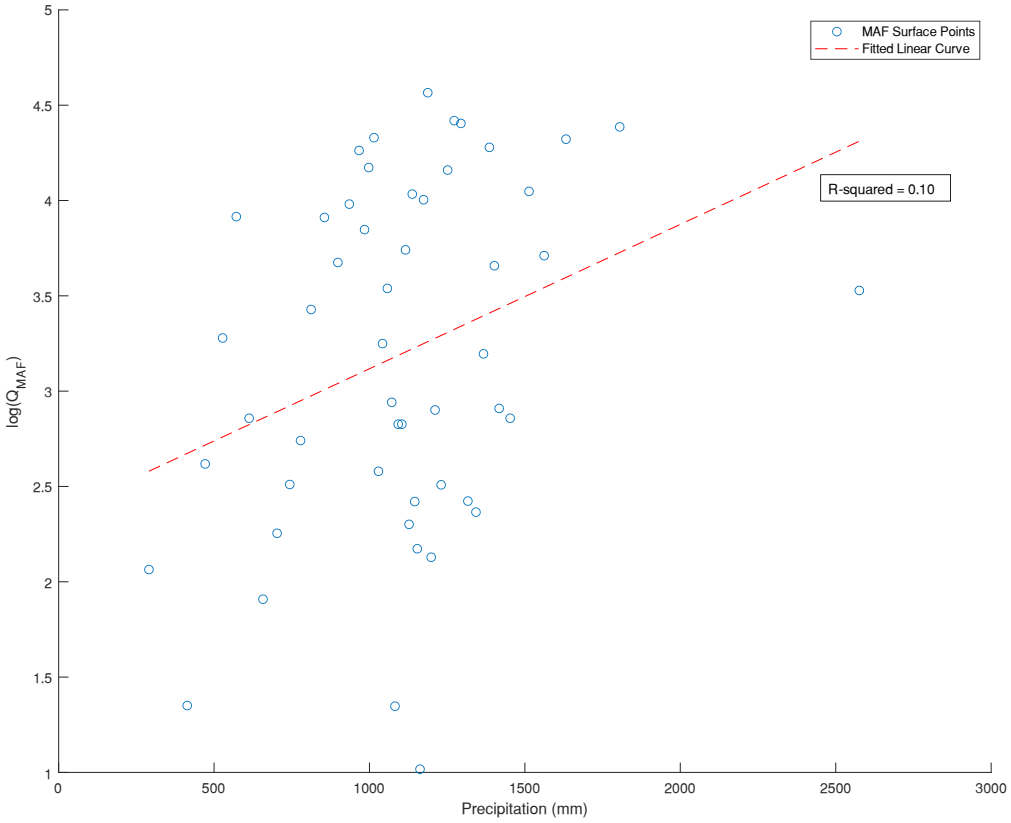


Figure 7. 2 - Example of how relationship is derived between MAF and precipitation.

Also necessary is the derivation of the growth curves. Within each climate zone, these are based on area and rainfall, using a hierarchical clustering algorithm to determine clusters of catchments, based on the distance between the catchments in Euclidean space, with the values of log(AREA) and log(rain) being used as coordinates.

Once a group of catchments has been clustered, the regionalisation technique of "swapping space for time" (Hrachowitz et al., 2013) is used, where the time series of each clustered catchment is joined together and normalised. By creating a longer time series from different catchments, the calculations of extreme values are deemed more reliable, as extra information has been gathered from many more catchments. Hence the swapping of space for time. As a result, each cluster has a longer and statistically more reliable time series from which to determine the 100-year growth factors. These curves are hence assigned to each cluster at a calculated centroid of the cluster area. Despite the potential which this concept

offers, it should nonetheless be treated cautiously, as it requires the assumption that series taken from different regions can be concatenated without modifying their statistical properties.

From the above derivation, this means that for any new ungauged catchments, if the rainfall, area and Koppen Geiger climate zone are known, it is possible to estimate its MAF from the derived surfaces and also estimate the growth factor from the derived growth curves.

7.2.2.4. Deconstructing the current rFFA method

While ostensibly providing peak event Q_{100} values for all catchments globally, there are a series of methodological assumptions which could be contributing to poor estimation of peak Q, as has already been seen in the preceding chapter. It is also important to break down these assumptions in the approach so that it is possible to accurately quantify the uncertainties and calculate the uncertainty bounds. The purpose of this section is to explain these assumptions and how these are expressed in the rFFA methodology.

Once the data is partitioned into each climate zone, it is further separated into bins according to the total precipitation. The data is split into groups of 50 gauges, an arbitrary value. This value is needed for each linear fit, as it ensures that there are enough "homogeneous" gauges to derive a reliable relationship of Q_{MAF} to area. For example, this means that for the GRDC gauges within the temperate climate, of which there are approximately 2600, there are 52 groups distinguished and separately fitted. Each group of 50 is then used to derive a relationship between area and MAF assumed to be a power relationship, as shown in figure 7.1. Once the fit is optimised and the coefficients estimated, these are used to create an interpolated surface across a range of area values.

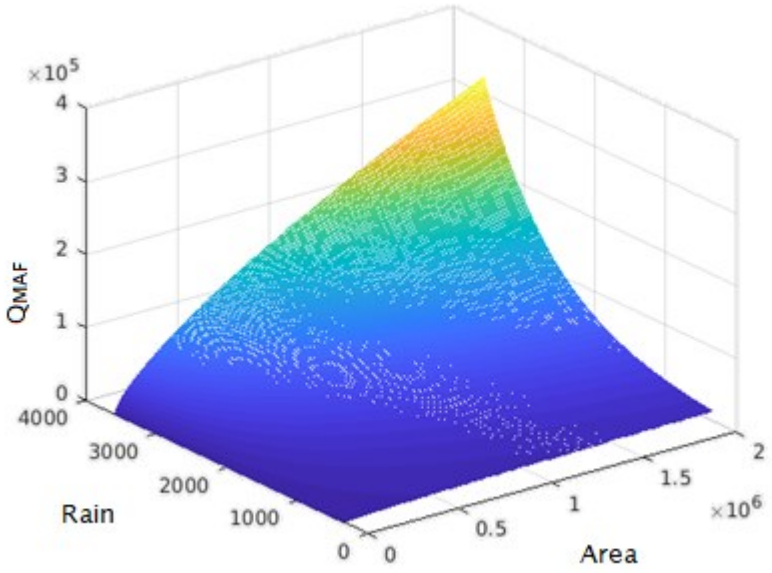


Figure 7. 3 - Final MAF surface product of the original rFFA method, with the positive relationships of area and precipitation to MAF achieved.

A key assumption of this method is that with increasing precipitation and increasing area, one always has increasing MAF, when measured in m^3/s. While generally true and not unreasonable, this does mean that the use of bins to create power laws means that this assumption is not adhered to. As each power fit occurs within a precipitation bin, there is no mechanism to prevent Q_{MAF} decreasing with precipitation, despite it seeming unlikely or counterintuitive. Therefore, the interpolated surface has to be smoothed with respect to precipitation. This represents a potentially major flaw in the model. The interpolated surface no longer represents actual values from the datasets, it is merely derived from them using parameters. Hence using a polynomial fit over the interpolants represents means that the surface shown in figure 7.3 is twice separated from the underlying gauging data. Hence, rather than the surface being the best estimate of a cloud of data points within the same space, it is instead a representation of a representation of the data, which is highly problematic. This is further compounded by poor R^2 coefficients, as shown in figure 7.2, which leaves the smoothed surface as a poor representation of the relationship of MAF to the rainfall and upstream area.

The derivation of the growth factor for an ungauged basin also has methodological assumptions to explore. As explained, an ungauged basin's nearest growth curve is identified

based on the predictors of log(AREA) and log(RAIN). However, rather than simply using the nearest centroid, the method stipulates the use of a distance-based weighting measure, meaning that the growth factors from every centroid are used, rather than the nearest one. In effect, the individual differences of growth curves are dampened due to this weighted averaging. This is a self-contradiction, in that the point of clustering is to identify characteristically different groupings of catchments, based on discrete difference in behaviour – such as small catchments with a peaky discharge against larger catchments with attenuated discharge (Blöschl et al., 2013). If the growth factors of the centroids are mixed together, then the characteristic differences are lost, and the method might lose its skill of identifying appropriate growth curves for a given catchment area.

Taken together, these flaws demonstrate some possible reasons for the poor skill of the method currently beyond that which is inherent in trying to estimate something so complex from a few variables. Certainly, the problem of estimating peak flows in a globally consistent way is hard enough without compounding through poor implementation. As a result, each segment of the method needs to be assessed and have its uncertainty individually quantified, before the overall uncertainty of a given estimate can be calculated.

7.2.2.5. Calculation of prediction bounds on Q100 estimates

From the above processes, we can calculate Q_{100} values. However, it is self-evident that this process is far from deterministic, and each step of the process has uncertainties which are not represented by a singular value, but rather should have a prediction interval – the range of values inside which we are confident the true value is, given the assumptions of the method and limitations on prediction by using only 3 variables. Therefore, we need to find each source of uncertainty and propagate it throughout the method. Each step of the rFFA methodology has an associated uncertainty. The power and polynomial relationships derived above each have prediction bounds associated with them, using the statistical information about the fit, such as goodness of fit and degrees of freedom. By their calculation, we are able produce probabilistic inputs to the next step of the process, and therefore propagate the uncertainty in a step-by-step approach.

We are interested in calculating non-simultaneous prediction bounds for the fits to our observations. This means that further predictions have a prediction interval associated with them, also known as the range of expected values given the predictor (i.e. area or

precipitation), and non-simultaneous means that each station input is treated as independent, with a singular value. For a given station *x*, the calculation of confidence intervals Q_{ci} are the following:

$$Q_{ci} = Q_{pred} \pm t(s^2 + \boldsymbol{x}\boldsymbol{S}\boldsymbol{x}^T)^{1/2}$$

Where Q_{pred} is the deterministic prediction of Q_{100}, and t is computed by calculating the inverse of Student's t cumulative distribution function, using our chosen confidence level of 95%. s^2 is the mean squared error, S is the covariance matrix of the coefficient estimates, given by:

$$S = (\boldsymbol{X}^T\boldsymbol{X})^{-1} \cdot s^2$$

Where X are the coefficients of our fits. Finally, $\boldsymbol{x}$ is the Jacobian evaluated at a specified predictor value.

To propagate the uncertainty of the power fit to the polynomial fit, a distribution must be assumed for the variance in the prediction. This means values from the distribution can be sampled, so that multiple fits can be made in the next step of the process. This is assumed to be a normal distribution. This must be done as the polynomial fits of the next step are based on the curve fit itself as opposed to the actual values, meaning that sets of possible fitted surfaces must be themselves fitted. These candidate fits are sampled by taking the mean and standard deviation of the fit distributions, taken to be normal:

$$\mu_{Q,x} = \left(Q_{ci,upp} + Q_{ci,low}\right) \div 2$$

$$\sigma_{Q,x} = \left(Q_{ci,upp} - Q_{ci,low}\right) \div 4$$

Following this two-step procedure for the calculation of the fits, we are left with distributions of upper and lower bounds for the final smoothed MAF surface, which links the MAF to the rainfall and area. Taking the mean and standard deviation of these sampled distributions, we are left with a single distribution of mean and standard deviation values. Hence, we can compute the probability of each value being selected, with the use of the normal probability density function. The distribution of these probabilities is then lumped together into an overall probability density. Hence the quantiles for the 95% confidence can be derived. These quantiles are used as the upper and lower bounds of the MAF surface.

The growth curves are similarly fitted from a distribution, although in this case it is the generalised extreme value distribution. To calculate the uncertainty of these values, the uncertainty of the parameters is directly sampled. Like calculation of the prediction intervals as in the case of the above fits, the bounds are derived from an assumption of normality. Hence the prediction intervals are derived from the inverse normal distribution:

$$C_{ci} = N^{-1}(0.95, \mu_{pred}, \sigma_{pred})$$

Where 0.95 is the confidence interval being considered, μ_{pred} is equivalent to the deterministic estimate of the parameter, and σ_{pred} is the estimated standard deviation of the coefficient value, calculated from:

$$\sigma_{pred} = \left(diag(acov)\right)^{1/2}$$

Where acov is the inverse of Fisher's information matrix. The diagonal components represent the asymptotic variances of the parameter values, hence the square-root returns a standard deviation.

With the prediction intervals derived for each growth curve, the uncertainty of each individual cluster is then considered. Because the growth factor for a new prediction is a weighted average of every growth factor in the given climate, the uncertainty of each growth curve must be independently sampled.

By sampling across the various distributions of the fits described above, ranges for each of the inputs necessary for the calculation of Q_{100} are found. There are prediction intervals attached to both the MAF and the growth curves. These are collectively sampled from and quantiles can be subsequently computed, which represent the final, lumped Q_{100} prediction intervals.

7.3.2.6. Sampling strategy

To propagate the uncertainty of the various methods through to the final desired output, each section must be sampled independently. This sampling is done through a Monte Carlo method, to achieve the desired output through a "brute-force" sampling of the variable space. The first fit to be sampled are the power relations. The number of power relations which can be sampled is dependent on the number of stations present within the given climate zone – remember that the number of stations varies extensively, and there is a

minimum number of stations permitted to each bin. Hence for tropical climate, with ~150 stations, 3 power relations are derived whereas for temperate climate, the existence of ~2600 stations mean that 52 power relations are derived and are therefore independently sampled from.

In an initial sampling of the temperate zone, 10,000 samples were evaluated, and a bootstrapping approach was used to establish that the prediction bounds had converged. From this, it was concluded that 10,000 samples would be appropriate for all of the climate zones. Using a Latin hypercube sampling strategy, a matrix of *M* by 10,000 samples was produced, with *M* being the number of power relations derived per climate zone. As the mean and standard deviation of each curve fit is already known, the sample is taken to be:

$$s_{m,n} = N(0,1)$$

Where $s_{m,n}$ is the *n*th normal sample of mean = 0 and standard deviation = 1, for bin *m*. The sample is then re-scaled to the correct curve through the use of the mean and standard deviations previously found:

$$curve_{m,n} = \mu_m^{curve} + s_{m,n} \cdot \sigma_{m,n}^{curve}$$

These curve samples carry through the polynomial fits, producing a distribution of surfaces which are smoothed. As the polynomial fits are used to smooth the surface, each sample surface has an associated prediction bound, meaning that from the distribution of surfaces, we end up with a distribution of prediction intervals.

Sampling for the growth curves proceeds in a similar method. Normal samples are produced in an M by N matrix, with M being the number of independent growth curves within the climate zone and N being the number of samples selected, again in this case 10,000. As the mean and standard deviation of the GEV parameters has been previously found, the sample values of each growth curve are found with:

$$\theta(\mu,\sigma,k)_{k,n} = \mu_k^{\theta(\mu,\sigma,k)} + s_{k,n} \cdot \sigma_k^{\theta(\mu,\sigma,k)}$$

Where $\vartheta(\mu,\sigma,k)_{k,n}$ is the GEV parameter set of growth curve *k* and sample *n*. Therefore, the sampled growth factors are found to be:

$$GF_{100}^{k,n} = GEV^{-1}(0.99, \theta(\mu, \sigma, k)_{k,n})$$

Where $GF_{100}^{k,n}$ is the growth factor for the 100-year event for the *n*th sample in the *k*th growth curve. These factors are then weighted together to get a final growth factor.

To reach the final distribution of Q_{100} samples, an extra variable is added to the sampling scheme of the growth curves to account for the uncertainty of the MAF. Hence the final sampling structure is actually of size K+1, meaning that the uncertainties of the MAF and the growth curves are independently sampled and then brought together into a joint distribution of uncertainty. As the upper and lower bounds of the MAF surface have been previously defined, these are then used as limits to define the normal distribution. Hence the sample of the MAF is taken from:

$$s_{MAF}^{n} = \mu_{MAF} + s_n \cdot \sigma_{MAF}$$

Hence for a given station *m*, the distribution of Q_{100} values produced are:

$$Q_{100}^{m,n} = s_{MAF}^{m,n} \cdot GF_{100}^{n}$$

Where $Q_{100}^{m,n}$ is the estimated Q_{100} flow for station *m* with sample *n*.

From this distribution of Q_{100}, the 95% quantiles are calculated and final prediction bounds are set.

7.2.2.7. Significance of calculating Q_{ex}

Although it is useful to have the prediction bounds on the Q_{100} estimate, it is worth considering that the effect of this uncertainty on the floodplain (the original reason for calculating uncertainty) will be attenuated by the size of the river channel. As discussed in chapter 3 and the other results chapters, the size of the river is actually a parameterised feature of model structure. The size of the river is determined by the bankfull discharge, which is itself an estimate. Therefore, if the bankfull discharge Q_{bf} has a return period of 5 years as opposed to 2 years, then one sees that the proportion of the peak flow which will spill onto the floodplain is reduced, as the capacity of the river will be larger. This remaining flow on the floodplain we can call the excess discharge, or Q_{ex}:

$$Q_{ex} = Q_{100} - Q_{bf}$$

There are two significant implications for the calculation of Q_{ex}. The first is that with Q_{ex} calculated, it is not necessary to re-run the model for different values of Q_{bf}. In practical terms, this would mean that much of the modelling chain has to be re-run, which is less efficient. Q_{ex} however would be functionally the same value as only estimating the floodplain flow, and the channel size becomes a less important factor under the assumption of completely full banks. As a result, it wouldn't be necessary to recalculate the bankfull flows under uncertain conditions, as the only flow of interest is the floodplain flow. Another significant factor is that it also has uncertainty attached to its calculation – it is dependent on the same MAF value of the Q_{100} estimate, as well as the uncertainty associated with its growth factor. It is therefore of interest whether the uncertainty of the estimates is reduced for Q_{ex} over Q_{100}.

7.2.2.8. Training and test sets

To validate the method, a split sample approach of the baseline dataset is being used. That means that the approximately 5,500 stations in the GRDC dataset must be split into training and test data. As the rFFA methodology is data-driven, holding back too much of the data would penalise the method unfairly; however, it is still necessary to have a decent sample of points in each of the climate zones. As such, 603 points were randomly selected from the dataset. This number was reached by taking an initial pool of a random 15% of the data, assumed to be enough to adequately test each method, assuming that some would be rejected as inappropriate for testing. A small quality control was placed on these values, to ensure their time series were appropriate for the analysis. The criterion for their selection was that the corresponding AMAX time series has more than 50 years of data, so that the observational Q_{100} has a reduced uncertainty.

Within the pool of test data, there is only 1 polar station. As there are only 15 polar stations in the entire data-set, we are not able to adequately identify the uncertainty and it must be omitted from any further analysis. It is also difficult to accurately assess the tropical climate zone, which has 3 stations within the test data, and only 192 stations in the training data. Arid, temperate and continental climate zones have a denser gauge network, meaning they are better suited for this data-driven method and we would expect the estimates of peak discharge to be more accurate and less uncertain, as was evidenced in the paper of Smith et al., 2015.

No. of gauges	Tropical	Arid	Temperate	Continental
Testing	3	70	300	229
Training	192	531	2301	1669

Table 7. 1- Numbers of gauges in each KP climate zone, for training and testing. Note the small number of tropical gauges relative to training set is due to the requirement of 50 years' worth of data for generating Q_{100}.

7.2.3. Results

7.2.3.1. Uncertainty bounds of old method

Figure 7.4 shows the uncertainty bounds of the test data, split up according their corresponding Koppen Geiger climate zone. It is worth noting here that the Koppen Geiger representation of the current method is the coarser version, and the definition of the smaller Koppen-Geiger subsets is not used. Overall, the definition of uncertainty in these plots is unsatisfactory. This is because, as we have discussed before, the final smoothed surface is a model of a model. Remember that firstly Q_{maf} is related to area as a power law. 500 interpolated points are then taken along each curve, converted to log space, and considered as actual datapoints for the comparison of Q_{maf} with precipitation, with a linear fit. This procedure, along with other assumptions of the method previously discussed, which could as individual assumptions have merit, produce an incoherent relation of area and rain to Q_{maf}, and it is reflected in the uncertainty bounds. Added to this, the majority of the stations in the GRDC network here are actually from catchments which are of a similar size, in the range of 100 to 5000 km^2. Although the method of Smith et al., 2015 ostensibly covers catchments of up to 2x 10^6 km^2, in truth there are no catchments of this size in the dataset. Note as well that it is simply not possible to quantify the uncertainty at every point, because some of the catchments extend beyond the set of interpolated points specified in the method, in terms or area or precipitation.

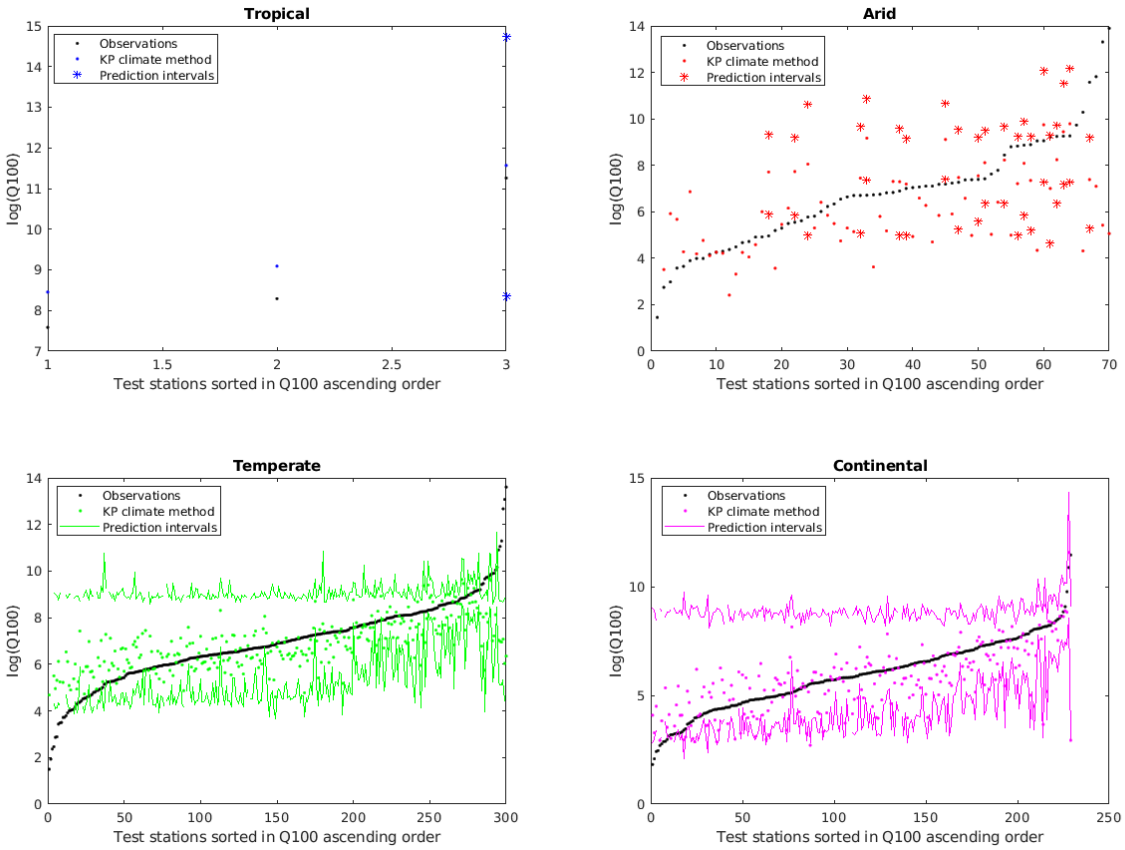


Figure 7. 4 - Predictions and uncertainty bounds of the 4 Koppen Geiger zones distinguished in the GRDC dataset. Note in the case of the tropical and arid climate zones, uncertainty bounds have been defined as points, as for some smaller locations, it is not possible to quantify the uncertainty, therefore there are only the predictions.

7.2.3.2. Deterministic predictions

Table 7.2 describes the error statistics of the original rFFA method in each of the climate zones. Taking the error of each gauge in the climate zone, one can derive the RMSE. However, given the range of catchment sizes, and therefore Q sizes across each climate zone, it is difficult to derive any conclusions based solely on RMSE. Instead, the relative RMSE normalises the error to the test Q_{100} value was calculated to represents the ratio of errors. From this we see that the arid zone is the most poorly predicted, followed by temperate and then continental. It is difficult to conclusively judge the effectiveness of the method in the tropical zone, as the relative RMSE is only based on three values, which in this instance are all very close together, with no outliers. Hence the reason that mean % error and R-RMSE show similar characteristics. However, purely on the comparison of mean % error with the other climate zones, tropical zones perform poorly. This matches with the previous findings of Smith et al., 2015.

Table 7. 2 - Summary statistics of rFFA performance for peak flow estimation. RMSE, relative RMSE and mean % error are considered, as the various magnitudes in each case mean the results show something different.

	Tropical	**Arid**	**Temperate**	**Continental**
RMSE (M^3/S)	$1.66 \cdot 10^4$	$1.52 \cdot 10^5$	$5.89 \cdot 10^4$	$7.38 \cdot 10^3$
Relative RMSE	1.08	4.32	2.98	2.14
Mean % error	99	113	74	41

7.3.4. Concluding remarks on results

The above results have shown that in certain cases, the rFFA method is able to predict peak flow with a basic level of effectiveness, as shown by the continental climate group, which has a mean % error of 41%. However, this is likely due to the density of the gauges within the continental zone, likely derived from US gauging data, as opposed to the tropical or arid zones which are more globally dispersed and have a much sparser network (see section 7.3.1.2). Meanwhile, the temperate zones also perform reasonably as many of the gauges are located in Europe (GRDC, 2018), defined as temperate for many of the gauges. By contrast, the discrete distribution of Koppen Geiger climates mean that there are a much more limited number of points for the tropical and arid zones. The use of the Koppen Geiger classification limits the information which can be derived from the data by deterministically assigning the climate of the stations in advance. As seen here, this means that some climate zones are sparsely populated while others are more densely populated with gauges. As a result of this variance in gauge density, we chose to look at alternative climate classification systems, to see if it would be possible to refine this aspect of the method, with all other variables being maintained.

7.3. Part 2: Refinement of the flood frequency approach using an alternative climate index and method

In this second part of the chapter, we will discuss a potential refinement of the rFFA approach and then compare the results of both methods to see what improvements we can leverage. As discussed in the previous section, there are several major methodological flaws in the

method, which we hypothesise are limiting the predictive skill of a regionalised flood frequency analysis, and are also obscuring the true predictive bounds of Q_{100}. By making a small number of relatively simple changes, the hypothesis is that it will be possible to elicit better predictions of peak flows for a given return period and the associated prediction bounds.

Much of the data will be the same – same area and rainfall data, same gauging stations, and the same split of training and evaluation data. What will change is how these datasets are analysed and whether improvements in spatial classification for the purposes of rFFA can be better extracted from them. As some of the previous results have shown, the use of Koppen-Geiger is too basic a classification system for our purposes and an alternative, finer-grained climate classification system has been selected. This system is the quantitative hydrological classification system of Knoben et al., 2018, which demonstrated better accuracy in estimating streamflow indices than the Koppen Geiger classification. In this context of fluvial flooding, this presents a satisfactory rationale for using it.

7.3.1. A refined regionalisation approach with more appropriate data

One of the most obvious flaws in the current set-up of the rFFA is the two-step fit procedure, resting on the notion that Q_{MAF} has a different relationship to the rain than it does to upstream area. While this is not a poor supposition, the execution of this is poorly formulated because each regression is performed independently of the other. As a result, the relationship of Q_{MAF} to precipitation is based upon previously made fits of Q_{MAF} to catchment area, rather than being specific to the data. This leads to a model of a model, as previously remarked. The consequence is that it is not really clear to what physical processes the regressions are linked or indeed how. Instead of relying upon two regression fits, a more elegant choice is to fit a surface of the rainfall and area data to Q_{MAF}. This can be done while still using the relations described previously:

$$Q_{MAF} = a.AREA^b$$

$$\log(Q_{MAF}) = c \times RAIN + d$$

These relations can be written as a single multi-variate surface equation, in the following form:

$$log(Q_{MAF}) = a \cdot \log(AREA)^b + c \cdot RAIN + d$$

By employing a regression of this form, it is ensured that the relationship between the data and the prediction is firmly maintained. By using a continuous form, the need to arbitrarily partition the data into bins is also removed, which has no physical basis and further obscures the value of the information available.

The major change we make in terms of the data used is to replace the Koppen Geiger climate classification scheme in favour of a quantitative hydrological classification system, first presented in Knoben et al., 2018. The new scheme uses 3 independent continuous indices to describe the global hydroclimate in terms of aridity, seasonality and fraction of precipitation as snow. By using continuous indices, the choice of how to partition data has to be made. This is done by using a kmeans clustering method on the data. This has some clear advantages over the use of Koppen-Geiger. For example, in figure 7.9 it is seen that with respect to the climate indices, the supposedly tropical stations are fully surrounded by temperate stations. By using a clustering method based on the climate indices, the different climate clusters are better defined.

The final difference made to the rFFA method is the calculation of a growth factor for a candidate/test station. Rather than calculating the centroids of each cluster, themselves being constructs of the data rather than the data itself, and then reducing the discriminative power of the clusters, the technique employed here will instead find an associated cluster for each test station, and will proceed with that growth curve alone, as described in chapter 3. While this may produce occasional mis-characterisations at the boundaries of each cluster, the overall effect will be that the characteristics of the station's catchment have full expression, and stations with peaky catchments have a correspondingly peaky response.

7.3.1.1. Description of hydroclimatic indices

The climate indices themselves are derived from global climatic information, collected from the University of East Anglia's Climatic Research Unit (CRU) (Harris et al., 2014). Monthly observations of precipitation, potential evapotranspiration (PET), and temperature have been compiled, for each 0.5° by 0.5° cell. To acquire indices relating to aridity and seasonality, Thornthwaite's moisture index (MI) (Willmott and Feddema, 1992) is used:

$$MI(t)\begin{cases} 1-\dfrac{E_p(t)}{P(t)}, & P(t) > E_p(t) \\ 0, & P(t) = E_p(t) \\ \dfrac{P(t)}{E_p(t)}-1, & P(t) < E_p(t) \end{cases}$$

Where *t* is a month within the representative year, and each condition determines whether that month is moisture limited or energy limited (Budyko, 1974). Using this moisture index, the average annual aridity and seasonality of a given cell is calculated thus:

$$I_m = \frac{1}{12}\sum_{t=1}^{t=12} MI(t)$$

$$I_{m,r} = \max\big(MI(1:12)\big) - \min\left(MI(1:12)\right)$$

I_m is the aridity index, which has a range of 1 to -1, representing the limits of humid, energy limited conditions and arid, water limited conditions respectively. $I_{m,r}$ represents the degree of seasonality within the chosen cell, as it measures the amount of intra-annual change in aridity within a given year. A value of 0 represents no seasonality, whereas a value of 2 represents a complete change of seasonality from fully humid (I_m = 1) to fully arid (I_m = -1).

The final metric used in this climate classification system is the fraction of precipitation which falls as snow, expressed as:

$$f_s = \frac{\sum P(T(t) \le T_0)}{\sum_{t=1}^{t=12} P(t)}$$

Which shows across how much time the temperature is below freezing when precipitation occurs.

These metrics are good representatives of the global hydroclimate, as they independently represent separate hydrological processes, such as total aridity and the seasonality of the given aridity, without being correlated. As a result, the value of one index is not a good predictor of another, indicating the indices represent distinct processes (Knoben et al., 2018). This means for the purposes of clustering stations for regionalisation, we can be confident that quantitative differences across these indices will translate to qualitative differences across the different stations. The figures below show the global extent of these metrics, first in terms of the various feasible combinations of the metrics and then the geographical positioning.

Figure 7.5 shows each possible combination of climate indices, calculated for the whole planet. Each point is a single 1-degree by 1-degree tile containing land. Within this tile, the 3 indices are calculated, and are then coloured according the primary colour palette. This colouring allows for the distinction of particular climate zones in figure 7.6. What is notable about this description of the hydroclimate is the relative sparsity of snowfall among the world's catchments. Also interesting is the 'U' shaped relation of aridity to seasonality. This makes sense, as when one is in a really dry or wet catchment, it could be expected to be like that the entire year, hence reducing the seasonality of that catchment.

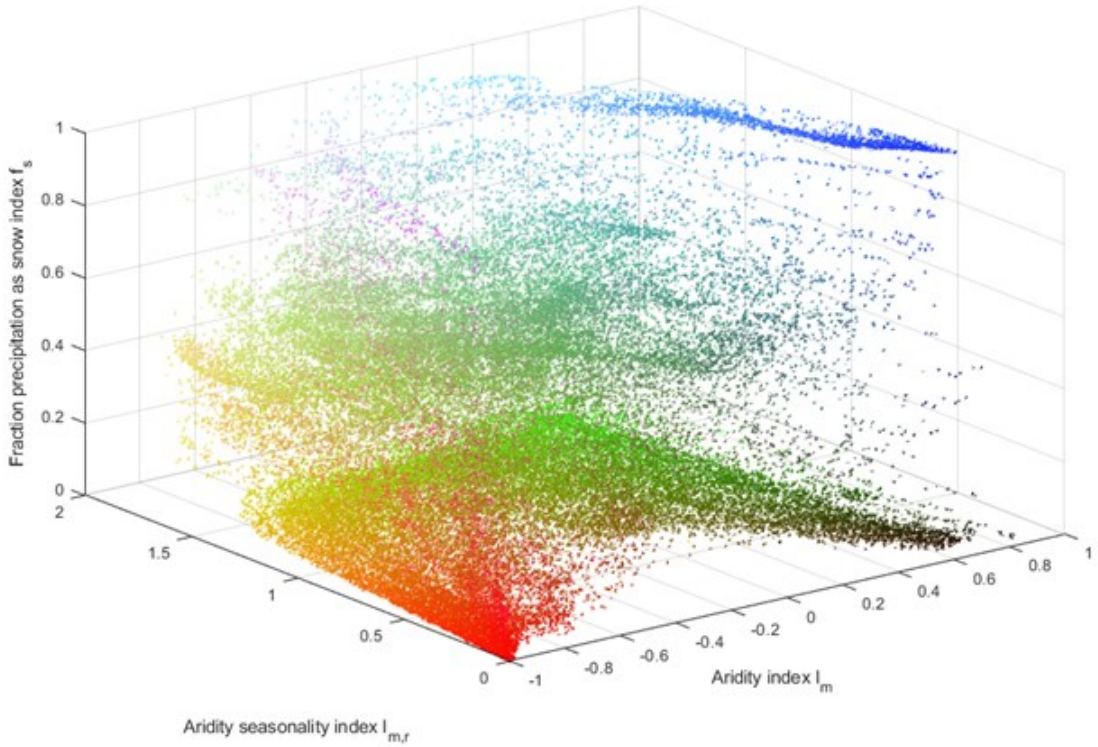


Figure 7. 5 - Each 1 degree by 1 degree land tile has been given coordinates and has its colour defined by the 3 independent indices of the new climatic system. Plot produced using methodology from Knoben et al., 2018.

As a sanity check, figure 7.6 conforms to expectation, as one can identify all of the major climate zones of the world, such as the hyper-arid Sahara desert, shown in red. Alongside this, all of the other desert regions of the planet are clearly defined – the Gobi desert, Australia, etc. Conversely, the mild, seasonality weather of Europe and the Eastern United States is

clearly defined as well, while the continental United States and Siberia are similarly coloured as well.

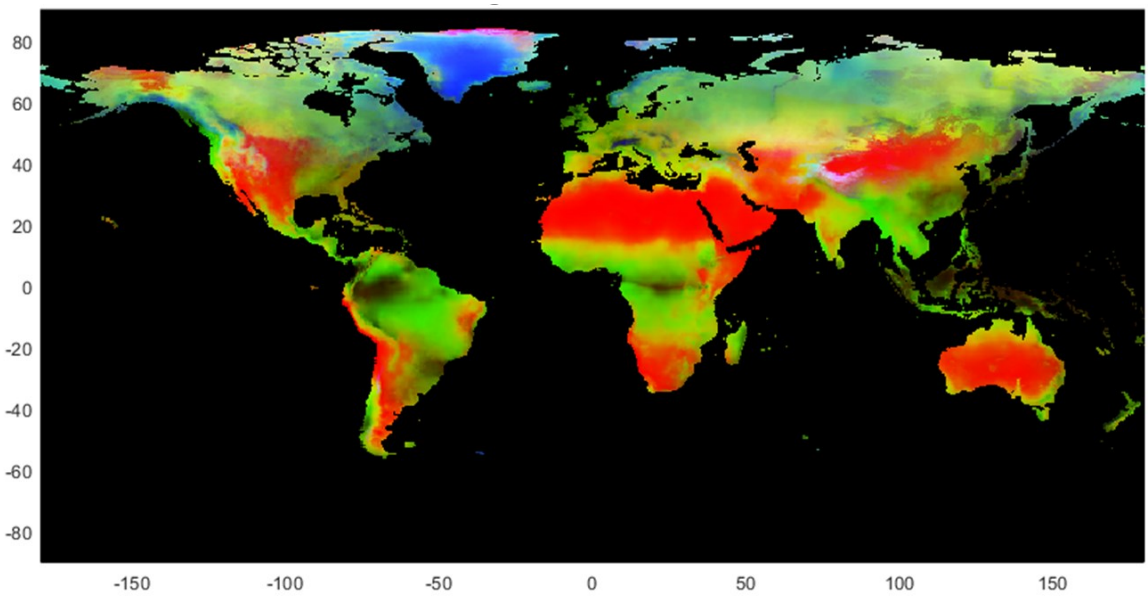


Figure 7. 6 - Projected climate zones onto the world. Plot produced using methodology from Knoben et al., 2019.

7.3.1.2. Intersection of GRDC stations with new climate scheme

While the climate indices shown in figures 7.5 and 7.6 represent the entire globe, the coverage of the GRDC data are more constrained. Figure 7.7 shows the coverage of the GRDC gauge network across the entire indexed climate space. What's immediately apparent is that the majority of the network is positioned around some very heavily gauged locations; mostly the United States and central Europe. As previously mentioned, this corresponds to the denser tropical and continental regions of the Koppen-Geiger classes. Beyond this, there are also many gauges across Indonesia and in Japan, while the middle East, Africa and South America are very poorly gauged – the arid climates of the Andean ranges are not represented at all. From this, it is expected that locations which are located near or around Europe and North America are those which are most likely to be successfully predicted.

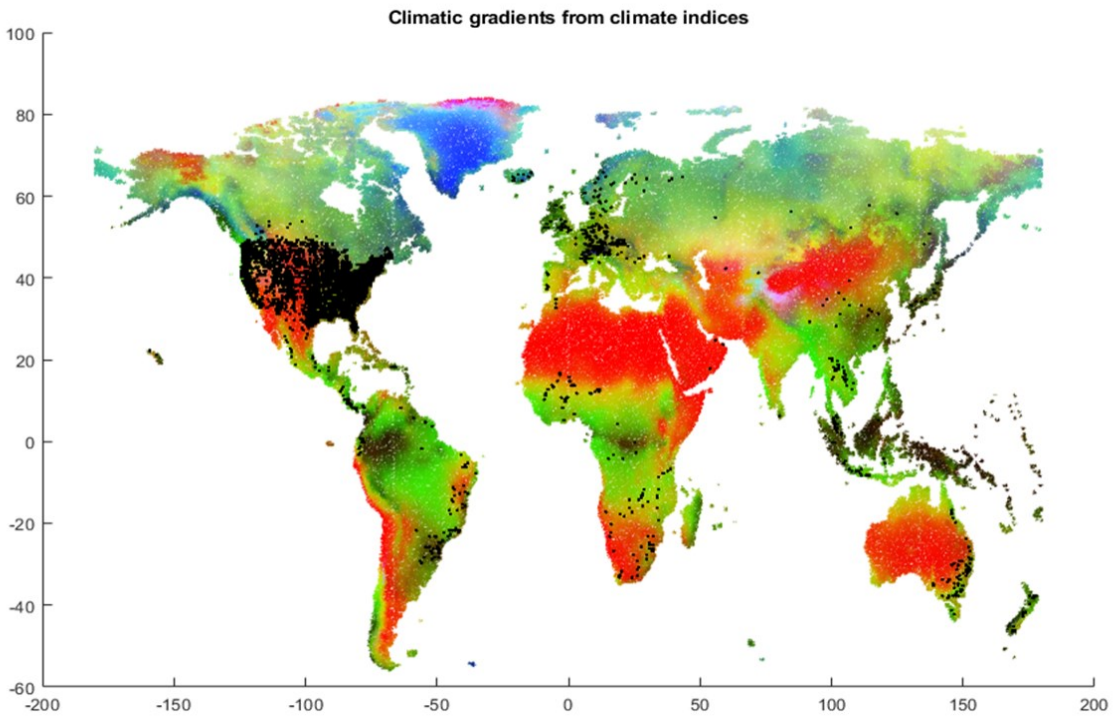


Figure 7. 7 - Position of the GRDC stations throughout the planet, and across the different climate zones.

By assessing the locations of the GRDC stations within the index space, it is apparent how climatically sparse the dataset is. Less than 10% of the climate index variability is represented in the GRDC dataset, as shown in figure 7.8. The effect of this will be that the number of clusters which are defined in the newer method are limited. Given that so much of the dataset will reside in the same index space, this means these areas are easily defined as their own climate zones. Areas that are more sparsely represented within the dataset, such as the Congo basin, are likely to be clustered with groups that are less well-defined.

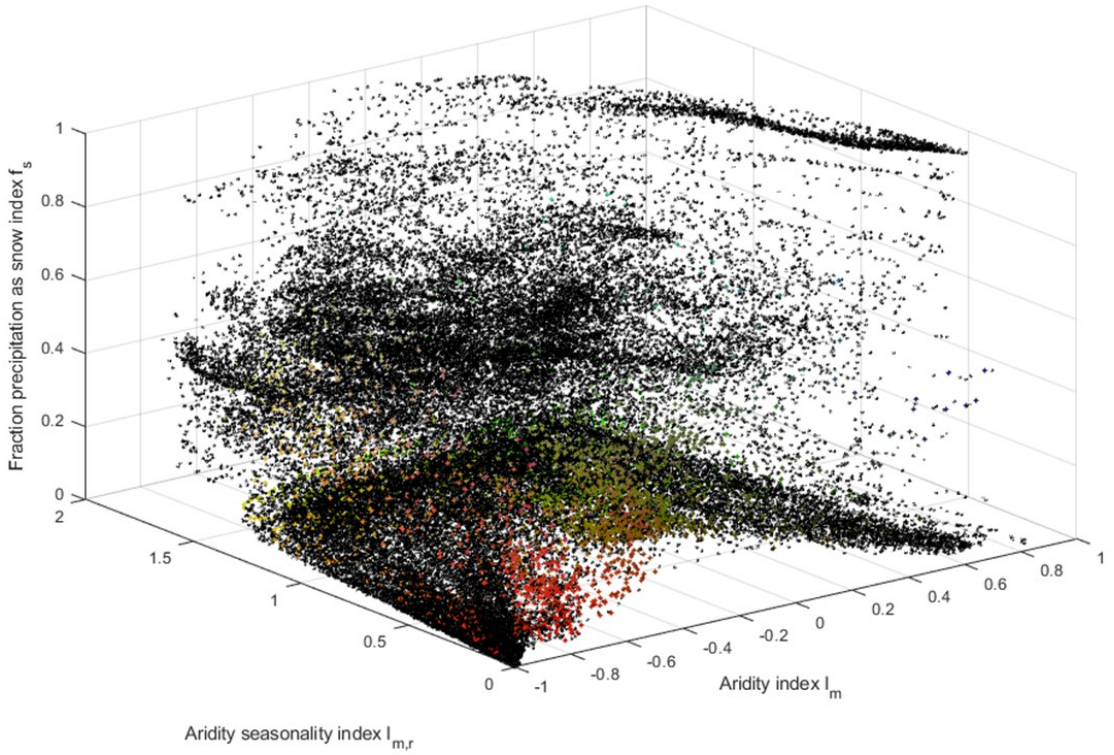


Figure 7. 8 - Climate variability of the GRDC gauging stations, with respect to the new climate classification. The black points represent the points which are not represented by the evaluation data.

We can highlight potential improvements of the new climate indexing approach over Koppen Geiger by projecting the Koppen Geiger classes onto the climatic indices. By doing so we can see some of the limitations of using the discrete Koppen Geiger classes. With reference to figure 7.9, it is clear that the Koppen Geiger classification system shares some similarity with the new system, as there is clear indication of a structure to the classes. It is also immediately clear is that the continental and temperate climate classes represent the vast majority of the points, leaving very few stations to the arid and tropical climate zones. The size of these clusters brings into question whether the clusters are suitable and could represent singular flood generating processes. This is further indicated by the status of the tropical climate, which is surrounded on all sides by the temperate class. This indicates that the designation of tropical in the Koppen-Geiger zones is based on some other factors than aridity and aridity seasonality, which may not be as pertinent to the hydroclimate. Given that the Koppen Geiger classification system is based on vegetation rather than hydrological features, this is likely a strong reason for the delineation.

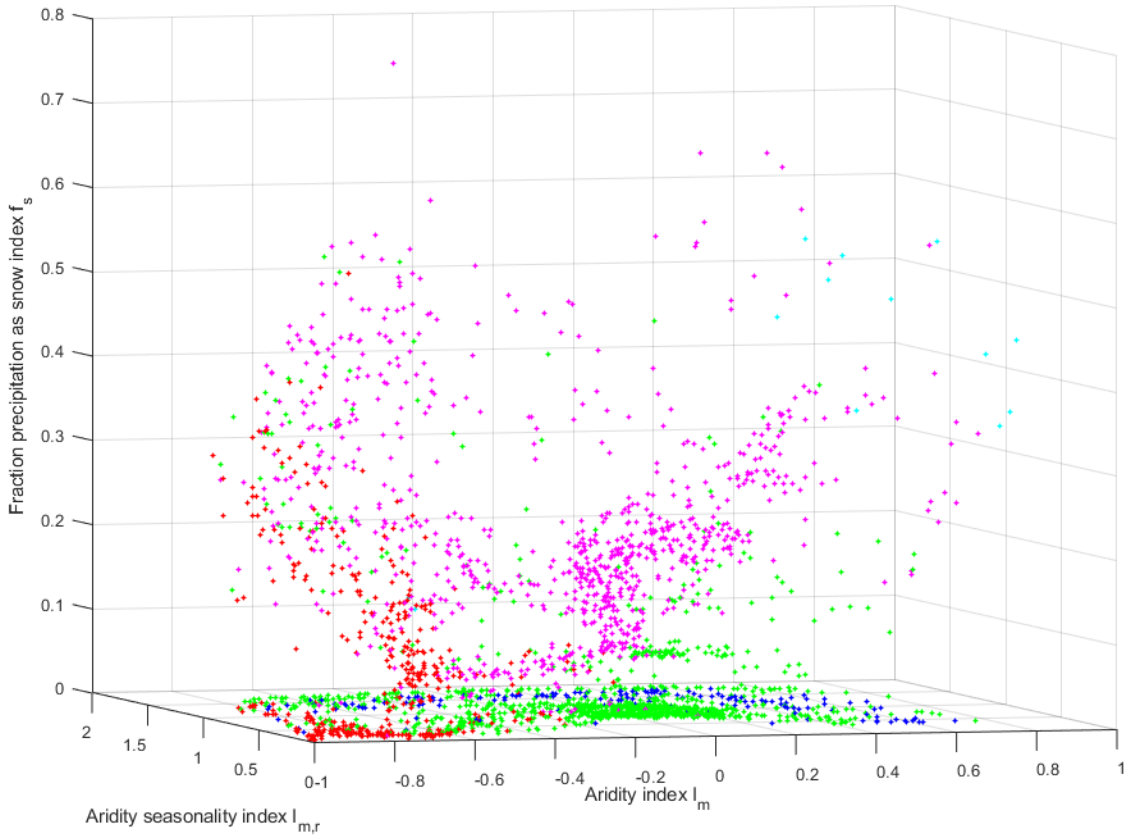


Figure 7. 9 - GRDC gauging stations coloured by Koppen Geiger zones, and coordinated by the quantitative hydrological classification.

Figure 7.9 also shows some of the limitations of each classification zone. The arid zone, while strongly correlated between the Koppen Geiger system and the new climate classification has some clear limitations as well. The demarcation of temperate and continental zones at the edges of the groups has some ambiguity. Another limitation of this zone and to a certain extent the temperate zones is that there is no demarcation between catchments with and without snow. Given that snowmelt is a significant mechanism across many catchments (Berghuijs et al., 2016), distinguishing the possible effects of this mechanism through regionalisation would likely be preferential.

7.3.2. New methodology for quantitative hydrological climate scheme

7.3.2.1. Clustering

As discussed, we are looking to replace the Koppen Geiger climate zones with a classification system which is more representative of hydrological mechanisms. The new indices employed are a suitable candidate for replacement, but as dimensionless numbers they exist on a continuum. While this is theoretically satisfying, it creates problems for regionalisation, and specifically how to determine an appropriate similarity across the gauging stations. Because of this, we are not able to avoid some discretisation of the climate space, although without the dictates of the Koppen-Geiger classes it might be possible to leverage a more effective regionalisation, as seen in Knoben et al., 2018.

The approach used here is a 3-tiered clustering hierarchy, to distinguish and then regionalise the data. The first tier aims to distinguish the stations by whether or not they have precipitation as snow. The reason for this is because in preliminary clustering, shown in figure 7.10, it was found that the effect of snow fraction wasn't being adequately accounted for. The reason for this is that the clustering approach works from Euclidean distances in space. As for approximately half of the catchments register as having no snow, these points are all a lot closer together in the index space. Thus, only aridity and seasonality were dominant factors in the clustering approach, despite our knowledge of flood mechanisms due to snow melt being potentially important. Therefore, to retain these effects, the dataset was split between catchments above and below 5% of precipitation as snow. This 5% is an assumption that a small amount of snowfall is unlikely to produce differences in flood generating mechanisms.

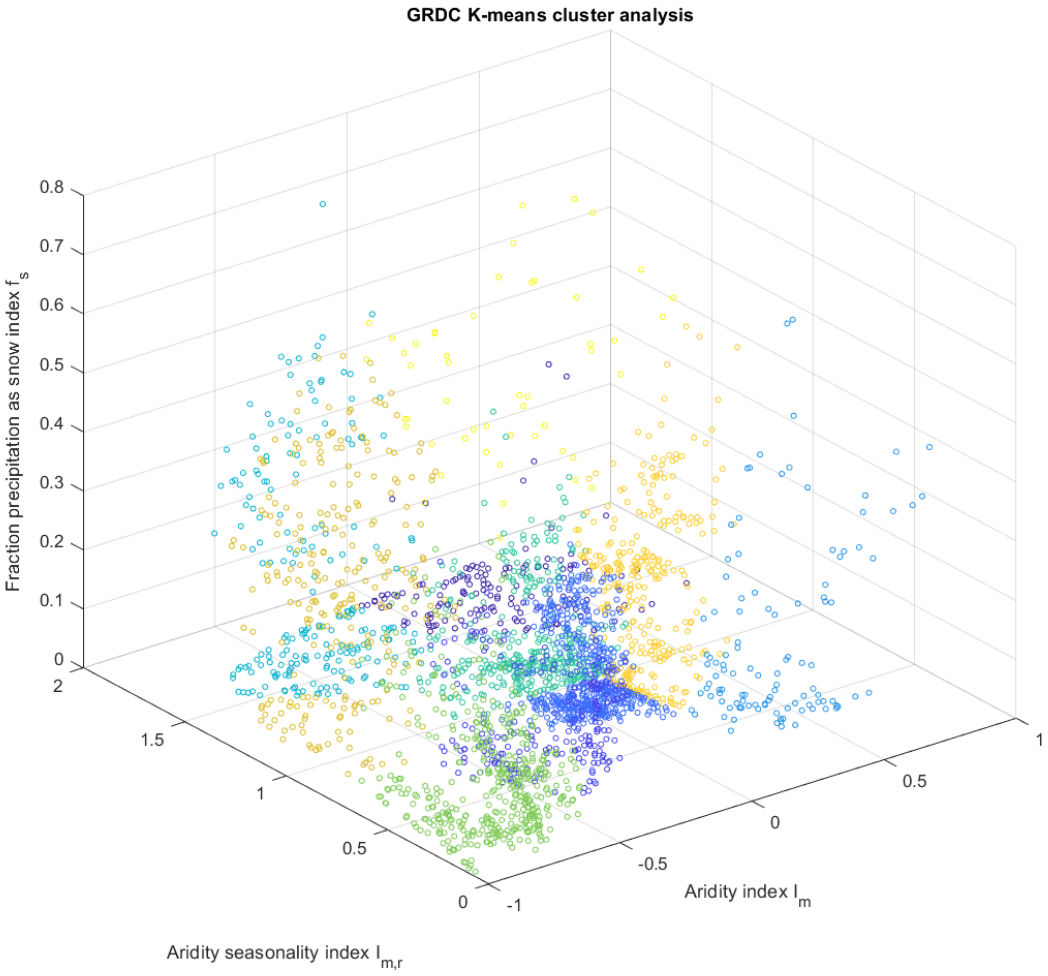


Figure 7. 10 - Preliminary clustering of the gauges - because of the density of gauges without snow, the snowfall index is barely weighted.

Once these two meta-clusters were distinguished, the same clustering approach with kmeans was used to partition the data into further hydro-climate zones. As it happens, the distinction of the snowfall meta-clusters splits the data into 2 sets of almost equal size. To further partition the data, the two criteria for clustering are that the clusters are satisfactory in terms of their membership mathematically, and also make sense scientifically. The mathematical criterion is driven by a "silhouette value", which given an indication of whether a given point is fully defined by one cluster or could quite easily be defined by another (Mathworks, 2020). Well defined clusters are therefore those which have high silhouette values for all the points. Figure 7.11 shows the silhouette values of the clustering scheme shown in figure 7.10. While satisfactory, we would still reject this clustering scheme as it does not account for conceptual

difference of precipitation as snow. As such, the clustering was undertaken separate for catchments with and without snow.

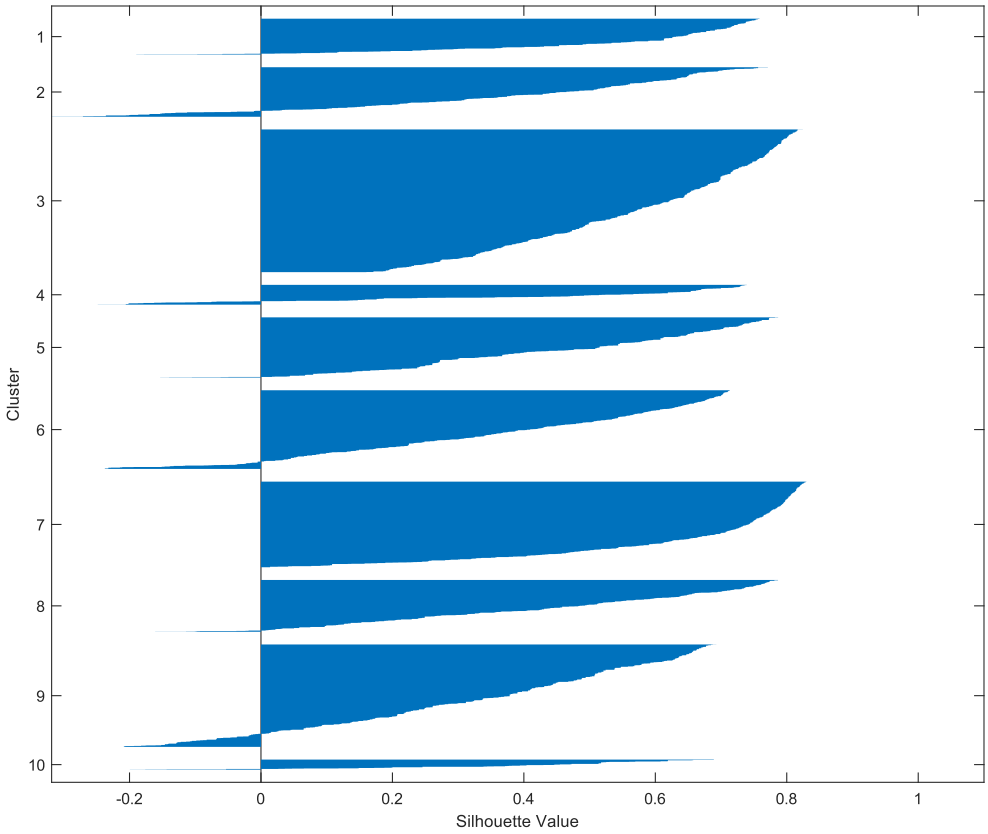


Figure 7. 11 - Degree of membership of points in each cluster: the silhouette value determines how well defined a point is by its cluster. A perfectly well-defined cluster has all of its points with a silhouette value of 1. Negative values occur when there are multiple clusters to which a point can belong. In practice this will never happen.

For snow catchments, it was found that 3 clusters defined the data in a meaningful way, as both the silhouette values were the best and they can be interpreted in a meteorological sense. These clusters can be described as approximately arid, semi-arid seasonal, and seasonal non-arid. While the opposite of arid is generally given to be tropical, in these cases where there is some snowmelt, such a designation would be inappropriate. The rainy catchments instead are most appropriately split into 4 clusters. Again, these clusters are well able to distinguish separate sections of the index space, and they make sense conceptually. For instance, there is a clear distinction of the arid non-seasonal stations from those which are more semi-arid and have an element of seasonality, perhaps representing stations in the

"Mediterranean" zone. Taken together, the 7 clusters of figure 7.12 are responsible for the calculation of MAF, as the data of the clusters is used for the regressions of area and rainfall against Q_{MAF}.

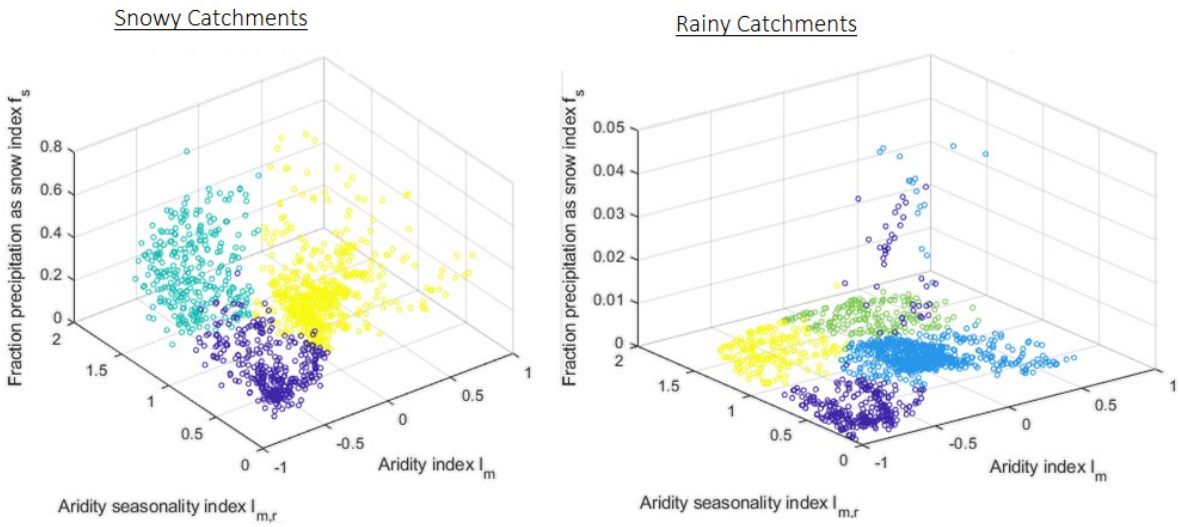


Figure 7. 12 - The 7 clusters using the new quantitative hydrological classification scheme.Plots distinguished based on whether or not they include a significant snow fraction, selected as f_s = 0.05.

So far, the clustering was done to distinguish climate zones in the data, which produce different MAF surfaces. This method is informed by Smith et al., 2015. However, there is still a need to produce growth curves. These are further partitions of the data within each climate zone, to account for the variability of catchment response within that zone. The result of this is the creation of a 3rd tier, which produces sub-clusters determining the sets of data to produce the growth curves. As correctly pointed out by the original method, sample size has a large effect on the growth curve characteristics and this representation should be retained in the new method as well. As such, the clustering algorithm derives sub-clusters based on area and rainfall of the stations within the climate zone. However, in practice, due to the much larger range of catchment area over rainfall (even within log-space), the sub-clusters derive the growth curves almost exclusively from the effects of catchment area.

7.3.2.2. Regression

Rather than using a series of univariate regressions, the new method uses a single stage multivariate regression, fitting a surface of Q_{MAF} to rainfall and area. In doing so, the uncertainty of the data to the fit is more easily communicated, instead of the need to sample across two fits and finding a lumped uncertainty. The least squares optimisation scheme is used to fit the surface, as described in chapter 3. An example of this fit is shown in figure 7.13,

computed in the curve-fitting toolbox of MATLAB. As the data is clustered across 7 groups according to their hydroclimatic indices, so there are 7 independent MAF surfaces, meaning that there are potentially 7 different response patterns of Q_{MAF} with respect to the predictor variables, because of the hydroclimate.

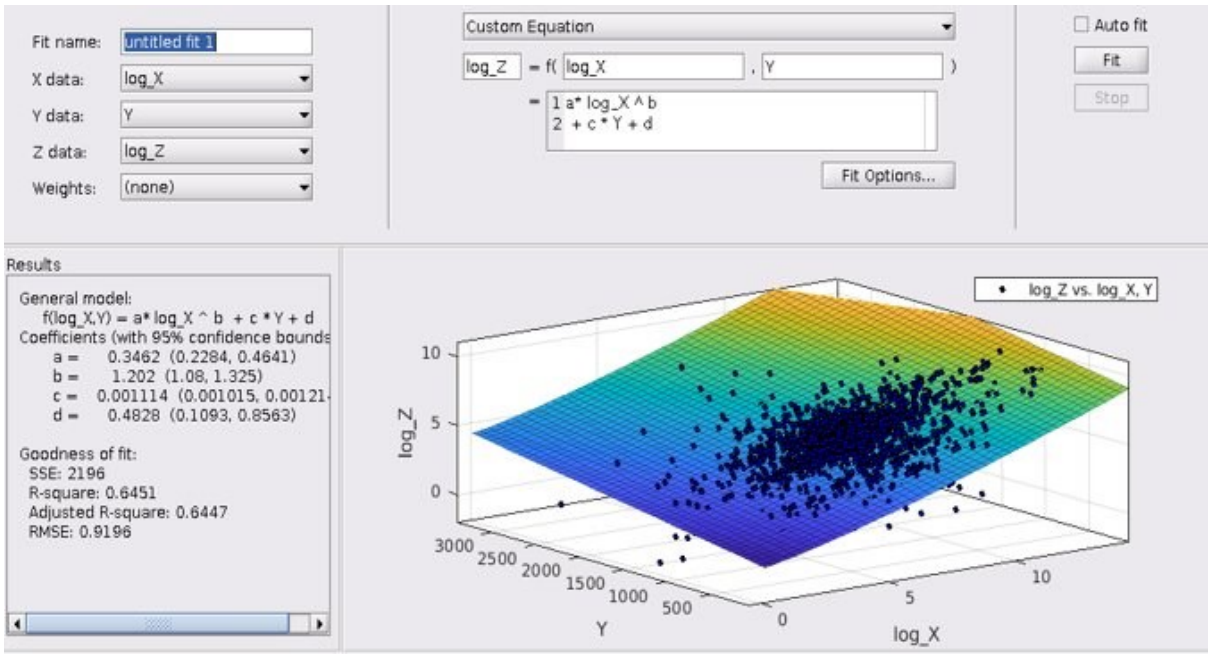


Figure 7. 13 - Fitting of the MAF surface with a model using the assumed relations of Smith et al., 2015.

7.3.2.3. Classification

Because the new indexing method uses continuous values instead of qualities as in Koppen-Geiger, each station must be assigned a climate type. The same is also true for the growth curves. Rather than using centroids, a simple classification scheme was employed. This is a decision-tree classification method (Sikorska et al., 2015). The training data is taken and assigned to each cluster. The algorithm learns the associations of each variable to each cluster, meaning that for the test set we can learn which cluster each station belongs to, meaning it is assigned to the most appropriate MAF surface and growth curve. In each case, the classifiers which were trained had a final success rate of over 99% in the classification, with errors only being committed at the borders of the clusters, where there is some ambiguity about precisely which climate the station is in.

7.3.3. Results

Having critiqued the old methodology and introduced the GRDC data datasets, this section now looks at comparisons of the old method against the new, in predictive capabilities and uncertainty estimation.

Table 7.3 below shows the summary statistics for the error of each method, previously seen in Table 7.2. At a first glance, it appears that the Koppen Geiger slightly edges the new method in terms of deterministic predictions, with regards to relative RMSE and mean % error. However, the newer method performs better with respect to the RMSE, although this value is limited due to the great variance of Q_{100} values. In essence, this table is telling us that one can expect more extreme errors from the Koppen Geiger method; however, the new method tends towards slightly larger errors of a moderate magnitude.

ALL GAUGES	**Koppen Geiger**	**Hydroclimatic indices**
RMSE	$6.68 \cdot 10^4$	$6.54 \cdot 10^4$
Relative RMSE	2.88	3.62
Mean % Error	65	76

Table 7. 3 - Summary error statistics of both methods, with respect to all the stations across all the GRDC stations.

Figure 7.14 highlights this conclusion, as one can see in the general error structure of both methods, there is a tendency for the Koppen Geiger method to contain more extreme deviations in error, while the HC index method has a more pronounced structure. In this plot, the sign of a perfect prediction is at the values of zero. What this graph is showing therefore is that there is a more structural error to the new method, as shown by the tendency to have a greater positive error – shown by the 'mean % error' statistic of Table 7.3. It is also clear that the very largest values of Q_{100} are poorly estimated by both methods.

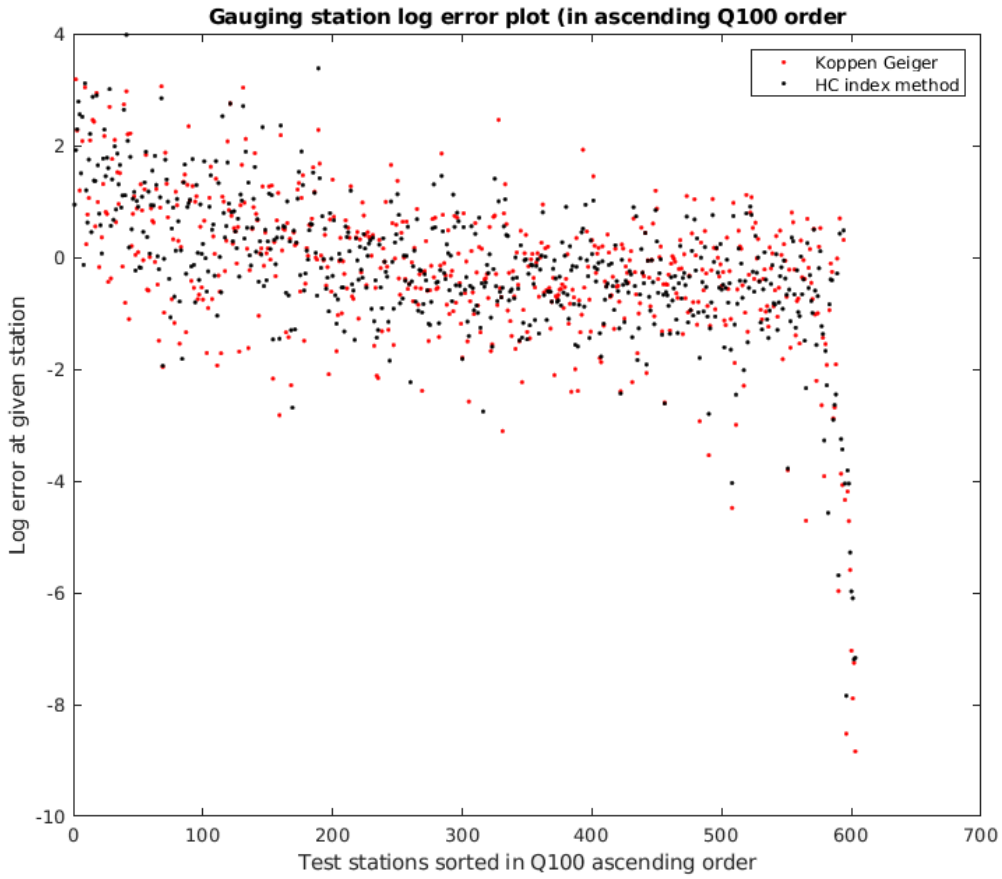


Figure 7. 14- Logged error profiles of old and new methods, showing a general trend of overprediction towards underprediction as the true Q_{100} increases. NOTE error is the log of the Q in m^3/s.

7.3.3.1. Comparison of uncertainty bounds for old and new method.

In figure 7.15, a full comparison of all the stations can be seen, old vs. new method. In red shows the old method and prediction bounds of the rFFA using the Koppen Geiger classification scheme, while the black points show the uncertainty derived from the new method. The blue line shows all the gauges taken as the point data, arranged in ascending order of size. It is also important to note that the y-axis is on a logscale, due to the varying sizes of the Q_{100} values. The new method (in black) shows a tracking of the Q100 uncertainty, indicating that the way the data is handled in the new method is more appropriate. However, it is clear that the size of the prediction bounds are still problematic in the new method, presenting difficulty in using it for reasonable estimation of peak flow. However, this is probably to be expected using only rainfall, area and a climatic classification. In general, the level of uncertainty stays the same across all the stations. The upper and lower bounds of the Q_{100} are approximately 8 times greater or lesser than the deterministic prediction. The

dominant control on this uncertainty are the MAF surfaces. Although they are better defined than in the case of the old method, the uncertainty bounds are still larger and indicate the need for more predictive variables. By contrast, the confidence bounds of the old method does not track the magnitude of the Q_{100} value as effectively. As already discussed, this is due to the limitations of the method, and the use of 2 linear fits, obfuscating the information value of the data.

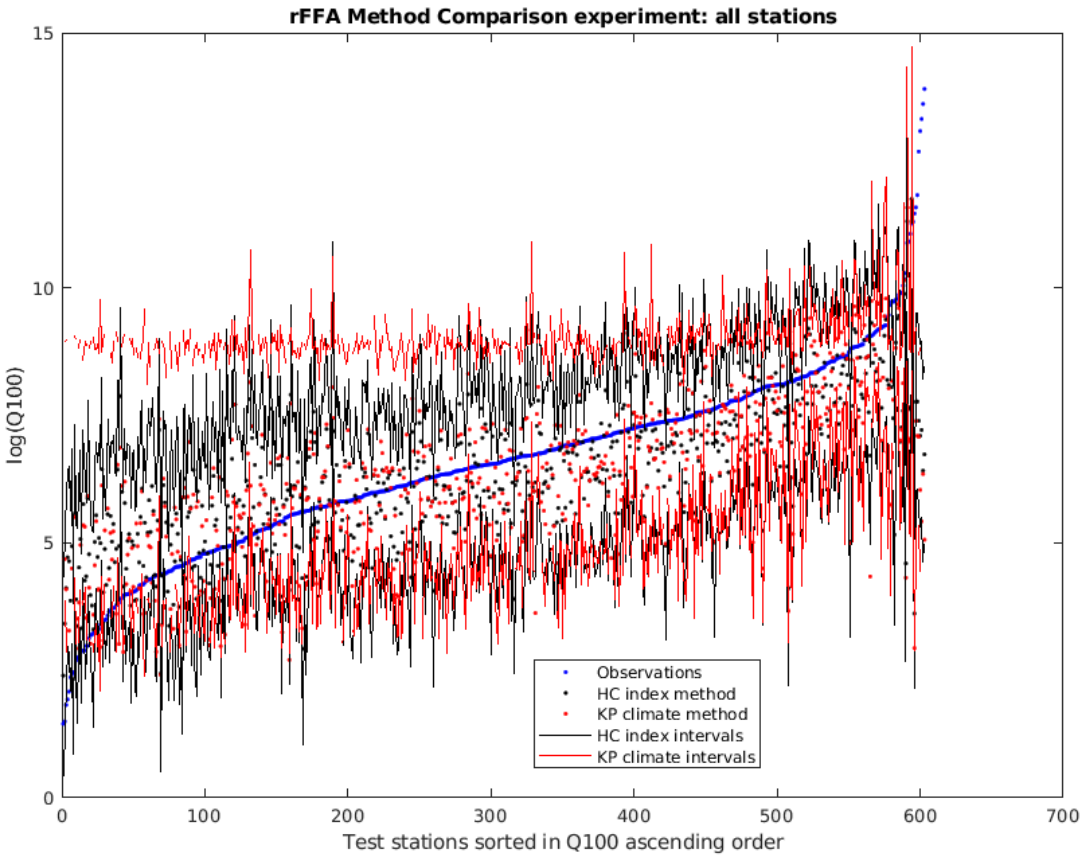


Figure 7. 15 - Predictions and prediction bounds of the old and new methods. The stations are ranked in ascending order by the known Q100 values.

7.3.3.2. Plots which show differences from the perspectives of each climate typology

While figure 7.15 shows the overall effectiveness of each method, it is important to look at the skill of each method relative to the climate zone, as this was the key difference made to the data. In figure 7.16, the skill of the methods in each Koppen Geiger climate zone is shown. In each case, the gauges of the new method come from newly defined climate zones; however, it is interesting to see if the predictions are better in terms of the Koppen Geiger

zones. In the case of the tropical climate zone, it is difficult to make definitive conclusions with only 3 gauges. Despite this, there is little difference in deterministic prediction. However, the uncertainty is much better defined in the new method. For particular catchments, the MAF uncertainty is unable to be calculated and this has happened here. Sometimes the partitioning of the data leaves too few independent gauges for the necessary analyses. But in the one example we do have enough gauges, the uncertainty is much smaller anyway. In the arid zone, again there is improvement in the number of gauges which have an uncertainty estimate, but where prediction bounds exist for both, there is inconsistency in which is better. This inconsistency will be a function of the clusters used at each gauge, as each surface fit regression will have predictions bounds of a different size.

In the case of the continental and tropical zones, we again see that the uncertainty is better defined, although the deterministic predictions - essentially a function of area and rainfall - vary little. Particularly as temperate and continental represent dense gauge areas, the corresponding climate clustering of the new method doesn't partition this data in a way that is significantly different, although some of the individual clustering of these areas will vary on the fringes. In these plots, a clear message of the new method's comparison to the old is that it doesn't tell us any more about what the peak flow is, but it does tell us more about what the peak flow isn't.

The clustering of the data according to the new method can be seen in figure 7.16. If the meaning of each individual climate index is considered, it can be seen how each of these clusters is well-defined. For example, wet and seasonal catchments shown in yellow can correspond to the tropical zones of Koppen-Geiger, although this zone is self-contained, unlike the tropical zone as shown in figure 7.9. The orange zone, mild in seasonality and wetness are indicative of locations like the temperate climate zones of Europe. Similarly, the zones of aridity for catchments with and without snow are sensible and clear. As mentioned previously, the similarity in predictive response for the old and new method will be as a result of the two densely populated clusters, shown here in orange and blue. These are areas of mild aridity and seasonality, with and without snowfall, and correspond to the densest sections of the corresponding KP climate zones of temperate and continental. The use of regressions with equivalent relationships in both cases then ensures that these dense areas have very similar predictive responses.

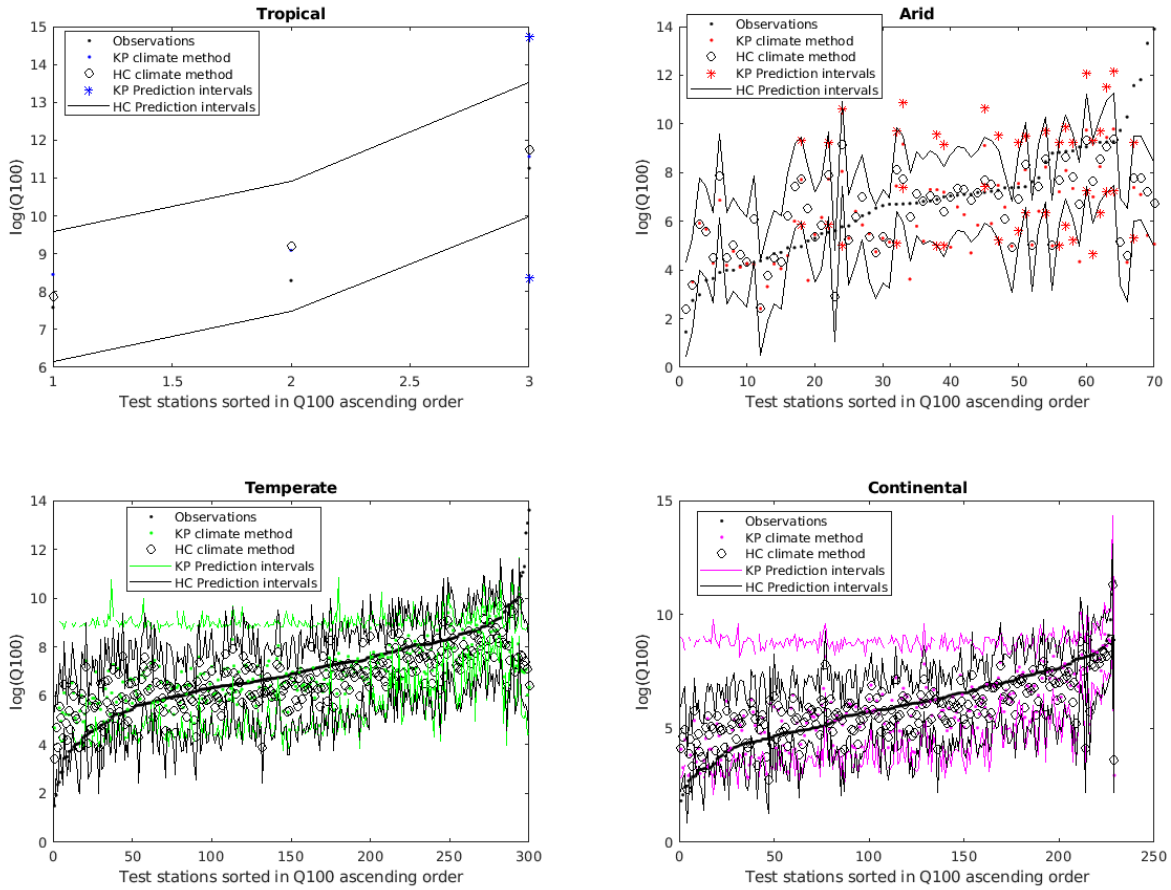


Figure 7. 16 - Performance of rFFA methods within each Koppen Geiger climate zone. In case of tropical and arid zones a sparsity of points means prediction bounds can't always been drawn.

Figure 7.17 below projects the results of the old vs. new comparison out into these newly defined climate zones, as figure 7.16 did with the Koppen Geiger zones. Each plot shows one of the clustered climate zones as defined in figure 7.16. In each plot are the accompanying predictions reached through the original Koppen Geiger method. In this way, it can be seen how each method compares in different climates, as viewed through each individual method. It is again important to note that the plots are arranged in ascending order of the size of the observed Q_{100} value.

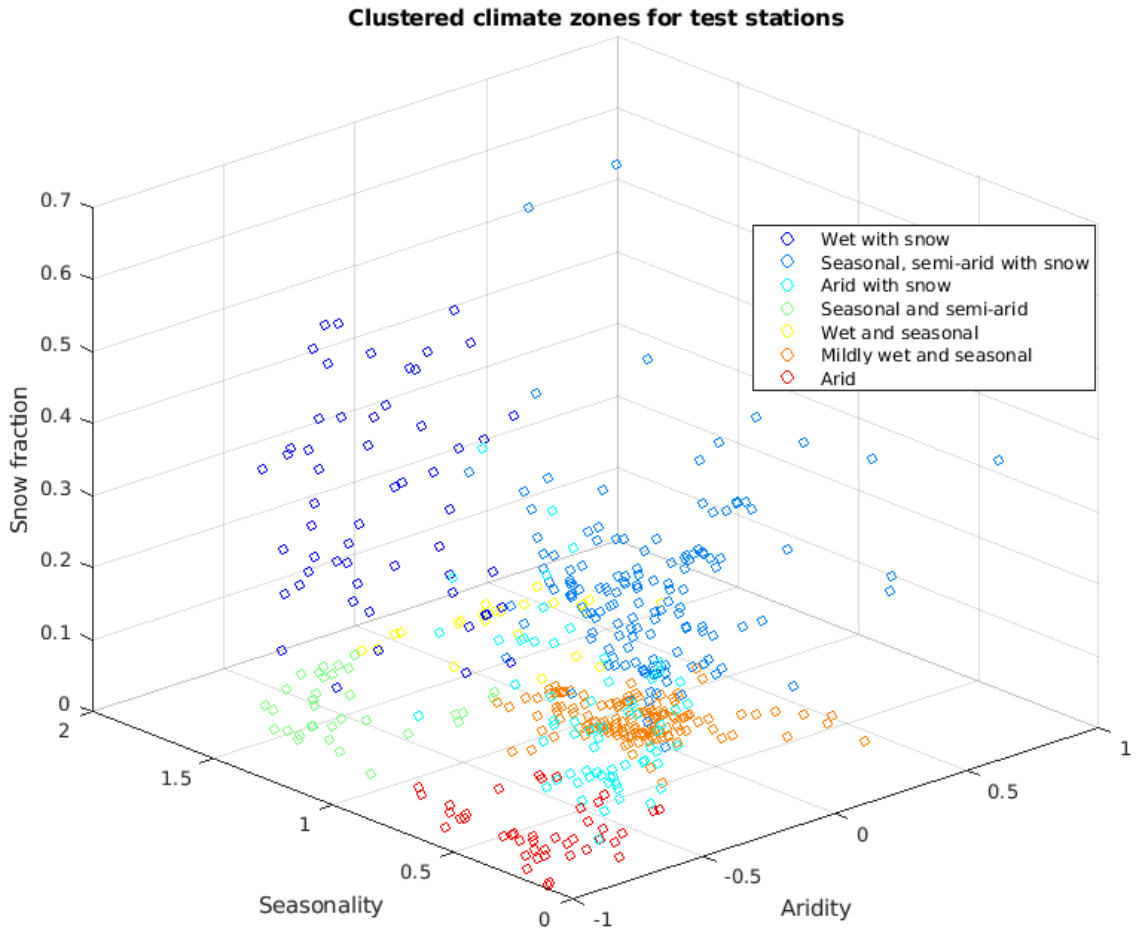


Figure 7. 17 - Clustered climate zones for the test GRDC gauges, using previous clustering algorithms.

In figure 7.18, it is again clear that the temperate and continental zones line up with two of the new climate zones. Further away from these areas, in other climate zones, the KP method becomes less discriminative and the new method shows clear improvement, and the uncertainty bounds reduce in size. This is in part likely to do with the vast majority of points being located in either tropical or continental. Interestingly, in "arid" and "seasonal and semi-arid" both methods fail to capture larger Q_{100} peaks. It seems likely this is due to the rainfall data used - average areal rainfall, whereas these events are likely the result of high intensity storm cells or something similar.

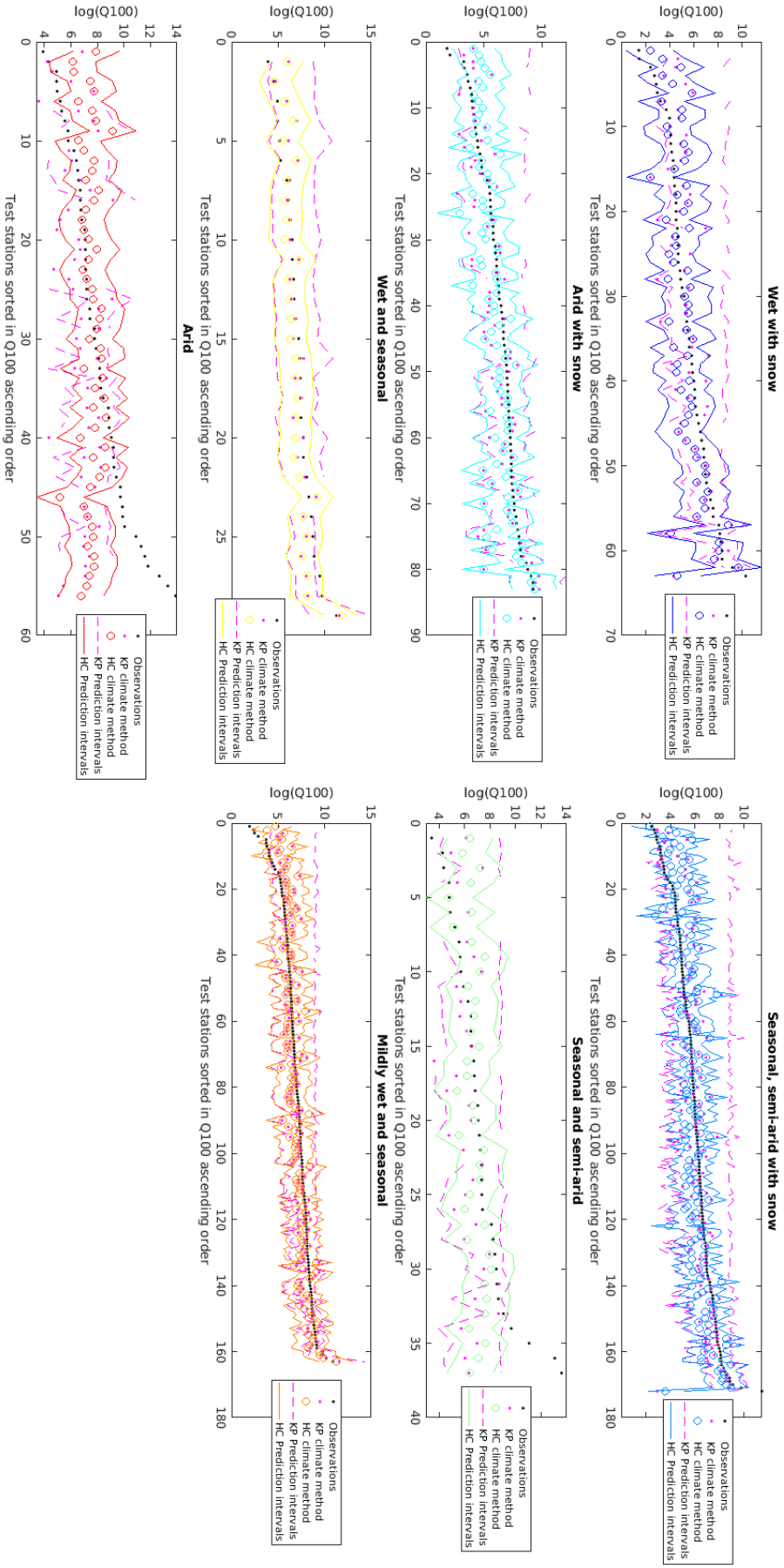


Figure 7. 18 - Comparison of each of the test stations with respect to the climate zones defined by the new method.

7.3.3.3. Comparison of Q_{ex}

Q_{ex} was assessed to look at the effects of flooding without having to account for the uncertainty of the bathymetry, as previously discussed. The following results show the predictions of Q_{ex} using the newer HC index method. Figure 7.19 shows a plot of Q_{ex} alongside the predictions of Q_{100}. It can be seen from the graph that the predictions of Q_{ex} are smaller than the Q_{100} values, as it accounts for the removal of the Q_{BF} value. Furthermore, as the interval of Q_{BF} increases, the prediction of Q_{EX} decreases. Another notable feature of this graph is that the Q_{EX} values have noise compared to the Q_{100} value. This is showing that these predictions all come from different collections of growth curves and MAF surfaces, meaning that the relation of Q_{100} to Q_{BF} differs in each case, resulting in a distribution of values below Q_{100}.

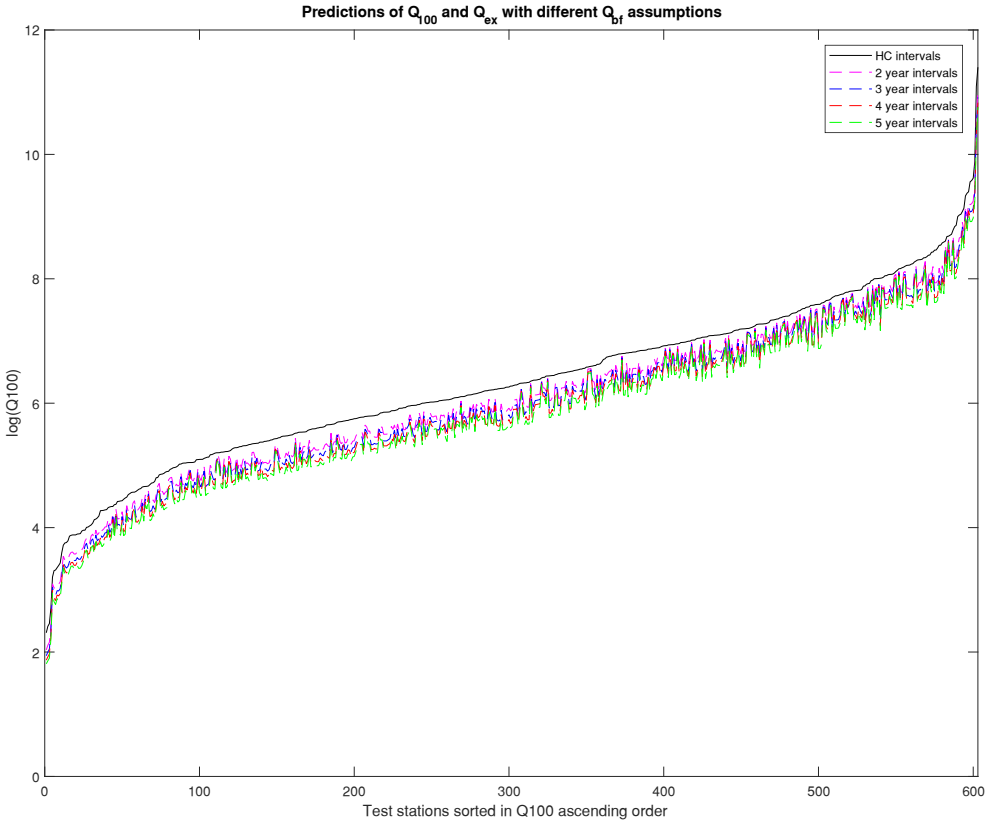


Figure 7. 19 - Plot of Qex for different values of Qbf alongside original Q100 prediction.

In the same way that the prediction bounds of Q_{100} can be derived from a large sample of Q_{100} estimates, the same is true for the Q_{EX} values. Hence in figure 7.20, the prediction bounds of the Q_{EX} values are shown with respect to those of Q_{100}. The uncertainty bounds have a smaller

value, corresponding to the smaller values of Q_{EX}. However, it is also noticeable that the range of the uncertainty bounds have reduced in size. On the upper bound, there is a greater drop in the value than on the corresponding lower ground, indicating that the estimated uncertainty is slightly less. Although it is not a significant reduction, it does show that by estimating Q_{EX} instead of Q_{100}, it is possible to reduce some of the uncertainty of calculations which is brought about by prediction of the channel bathymetry. This is significant in the context of the Bristol GFM and GFMs more generally. By focussing on only the Q_{EX} and its associated uncertainty, the modeller is able to more effectively parse the uncertainty of flood predictions. While this has been shown in the case of the Bristol GFM here, the limitations to measurement of bathymetry or discharge in remote sensing mean that other GFMs will face similar issues of uncertainty relating to the capacity of the river basins. For example, in the CaMa-Flood methodology, a return period channel geometry is assumed as in the case of the Bristol GFM (Yamazaki et al., 2011), while GLOFRIS assumes a conveyance capacity before determining how much water to propagate onto the floodplain (Winsemius et al., 2013). This effectively means that the derivation of Q_{EX} would also be of value in their frameworks as well and the inclusion of uncertainty bounds means that is it possible to propagate the uncertainty of inundation.

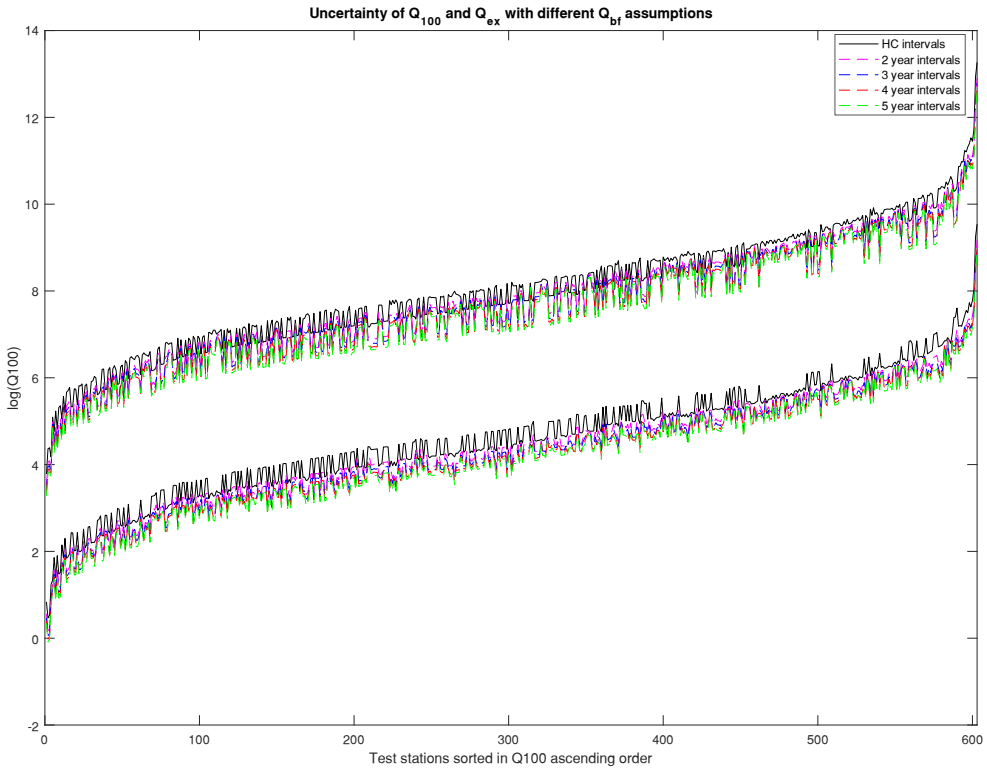


Figure 7. 20 - Prediction bounds for Q_{ex} given multiple Q_{bf} values alongside original Q100 prediction values.

7.4. Discussion

Both methods fail at capturing the extremities of the dataset. This will be due to the characteristics of the dataset itself, namely that the stations are not uniformly dense throughout the space. There are a very small number of stations with either low (<50m^3/s) or high (>50,000m^3/s) flows, which correspond to these outlier sections. As a result, the least-square regressions employed tend to ignore these points, focussing on optimising the fit in the middle of the much denser clouds of data. Added to this is the fact that most of the points within the GRDC dataset have similar catchment areas, meaning that in this context of the selected regression fits, both methods will struggle to represent the variability of the mean annual flood given just the area and rainfall.

An interesting development through the analysis is the changing shape of the MAF surfaces across each of the clustered climate zones. In the Koppen Geiger method, while there were different surfaces employed for each zone, the reality was that they all had the same characteristic shape, described by figure 7.3, informed by the handling of the data in the

method. The imperative to smooth the data, so that there is an increasing relationship with respect to both area and rainfall, is the dominant factor and mostly responsible for this uniformity. By contrast, the new method produces MAF surfaces which appear to be genuinely different in their characteristics, as seen by comparing figure 7.3 with figure 7.13. This is interesting, given that although the clustering algorithm itself is a purely mathematical entity, it has also contributed to a more realistic representation of MAF, at least in terms of area and rainfall.

It is also worth noting that the actual clusters themselves are a function of the GRDC dataset used, rather than being concrete specifications of climate type. As mentioned, the dataset used here gives a sparse representation of the global hydroclimate, instead representing a sample. Hypothetically speaking, the incorporation of more stations into the dataset, across a larger range of climatic indices would change the number and type of clusters used here. This can be viewed as a strength rather than a weakness, as it highlights a flexibility to the method, which balances the representation of climatic difference with the need for statistically significant groupings of gauging stations. The alternative would be something similar to the Koppen Geiger classification seen above, where there are either too few stations as in the case of the tropical zone, or there are so many that more climatic variation could be beneficial, as is the case with the temperate and continental zones.

Despite some of the clear improvements presented here, it is still apparent that much more work needs to be done on global regionalised flood frequency analysis. The reduction of the MAF surfaces to area and rainfall, conditioned by a climate typology, is unsatisfactory. When one considers the number of variables and parameterisations which can be present in a distributed hydrological model (Hundecha et al., 2016; Kauffeldt et al., 2016), it is clear why these variables alone are not enough. The process representation of the peak flows is highly reductionist. Effects of geology and landcover for example are not currently represented, which influence such hydrological signatures as base flow index (Hailegeorgis & Alfredsen, 2017a; Salinas et al., 2013). Despite there being a much larger pool of data from which to derive information (Gorelick et al., 2017; Lin et al., 2016b), none of these signatures are being effectively accounted for.

Another implication that can be taken from this work is that the datasets used are lacking in the most appropriate information. The use of an average rainfall dataset in the context of

events which are often characterised by extreme rainfall is quite likely a strong reason why the correlations of precipitation to MAF appear as weak, as often storm intensity is a more appropriate predictor variable for flood, which is very difficult to get correct in arid zones (Farquharson et al., 1992; Zaman et al., 2012). Perhaps a better approach would be to consider the rainfall as having multiple indices which take account of independent processes, like the new climate typology (Dawdy et al., 2012). In this sense, the average rain becomes a smaller component, alongside the intensity and seasonality, which better distinguishes rainfall events as having different meteorological components.

It is worth noting that in the analysis done here, assessment of model performance and the effect of uncertainty in the rFFA method were done using the "observations" of Q_{100}, which in essence assumes the value to be deterministic. However, as discussed in section 4.3.3., there are in fact large uncertainties present in the "observed" Q_{100} values, with deviations regularly exceeding 60%. This means that in much of the validation dataset, we could expect to see overlap of the confidence intervals in the predictions and the observations. However, given the complexity of the uncertainty quantification in just the method and the corresponding computational load this produced, the Q_{100} was treated as deterministic. If we compare the differences in the size of uncertainty, we see that they are not equal. The uncertainty of the rFFA is still much greater than the uncertainty in the observed Q_{100} values, given by the fact that the estimates have to be calculated in log space. However, it does mean that some of the uncertainty can also be apportioned to this observed uncertainty, and suggests that more work is needed to understand the relationship of the observed uncertainty to the estimation uncertainty. This means in effect that the sole reason for errors in either approach is not wholly due to the choice and accuracy of the hydroclimatic indices.

Through these examples, it is clear that we should advocate for a more complete approach to the regionalised flood frequency analysis. An example can be seen in the UK's flood estimation handbook, a regionalisation method which, through the wealth of information about UK's catchments, produces some highly complex data-driven relationships (Kjeldsen & Jones, 2007; WHS, 2014). The implementation of more rigorous relations with available data will have to be employed for the continuous improvement of a regionalised flood frequency analysis with some discriminative power. Area and rainfall are not enough.

7.5. Conclusion

This study has focussed on improving the data-driven methodology which underlies the regional flood frequency analysis of the Bristol global flood model. For the first time, the uncertainty of the flow predictions of the Bristol GFM have been quantified. It was found that due to a flawed methodological approach, the uncertainty estimates being produced were not a meaningful reflection of the actual uncertainty present in the prediction. As a result, changes were implemented to the design of the method, which more correctly took account of the uncertainties in modelling the data.

Within the methods used here, it has been shown that the dominant source of uncertainty for both the new and the old method is the calculation of Q_{MAF}. The implication of this is that there is more work to be done to address the lack of skill in this regard. Both methods have clearly demonstrated that while important, the area and average rainfall of a given station's catchment is not sufficient to provide an acceptable prediction, or confidence interval. These findings are concurrent with the practices of national scale methodologies, such as FEH, which advocate the use of a large number of explanatory variables for flow prediction.

The use of Koppen-Geiger climate zones was demonstrated to be insufficient for the task of extreme flow estimation. Given its focus on more general climatology, its applicability to the hydroclimate and hydrological signatures has been brought into question, as the discrete partitioning of climates is for purely hydrological purposes. Instead of assigning stations to a given climate in a wholly deterministic manner, the use of continuous climate indices gives more scope for difference between gauging stations and allows the user to determine their own degree of acceptable similarity between the stations. In the context of the current GRDC dataset, the differences of the response of MAF to area and rainfall has been demonstrated.

Changes were also made to the clustering and regression methods used, and a classifier has been added in the place of cluster centroids. These changes represent a more robust methodology which quantifies the uncertainty in a more sensible manner, as the relation of the model to the data is clearly understood, rather than being obfuscated by sequential processes. Beyond this, the methodology is now also more flexible with respect to using different datasets, as the clustering approach ensures that significant groupings of data are always used for a given regression. It would be possible to use the same principles and apply them to a different rFFA method.

From the work done here, we are now able to propagate uncertainties into the GFM boundary conditions, meaning that the uncertainty in GFM flood predictions can be examined within a Monte Carlo framework. This has further implications for the use of the GFM's in a risk-based approach, although the scales of the uncertainty do mean that more work is needed to find appropriate datasets to drive better predictions in the model, and reduce the overall uncertainty. Furthermore, the inclusion of Q_{ex} into the uncertainty framework means that these methods of calculating uncertainty can be translated to other GFM frameworks and address a core issue for inundation modelling which is so reliant on remote sensing.

7.6. Post-script

Throughout this chapter, the scope of improvements which could be made to the Bristol GFM was focussed on only the flow generating component, characterised by the regionalised flood frequency analysis. Although this is a small component of the overall modelling chain, the work of the preceding chapters has meant that we have identified this as the most important component, and improvements made here are pivotal to the success of the entire modelling framework.

This concludes the final results chapter of the thesis. From here, there are some clear further steps which could be done – namely, to begin a sensitivity analysis of the regionalised flood frequency analysis. The next step would be to do further modelling but make use of the new prediction bounds to produce an ensemble of hazard maps, and examine how the uncertainty has propagated through the model.

8. Discussion

Looking back from the beginning of the thesis to the final results chapter, the work was to begin with the aim of understanding the workings of the Bristol GFM in a more systematic way. Despite some initial successes with the global modelling methodology (Wing et al., 2017) and an extensive literature with which to validate the underlying hydrodynamic model LISFLOOD-FP (Aronica et al., 2002; Biancamaria et al., 2009; Komi et al., 2017; Sanyal et al., 2014), there was no clear description of exactly how the Bristol GFM works and which parameters influenced the outputs. This concern is illustrated by the currently sparse literature on GFM validation compiled by the community to date, with only two studies assessing the ability of multiple models to produce similar flood estimates; broadly speaking, this ability does not exist (Bernhofen et al., 2018; Trigg et al., 2016). Instead, what is mainly known is that each model has the individual means to produce estimates of hazard, and does so with broadly the same data streams as the others, but through different modelling strategies and implementations. This means that there is still the need for an overall validation framework for evaluating differences in the model outputs (Hoch & Trigg, 2019).

Hence it was necessary to begin the process of unravelling the various assumptions which underpin the global flood modelling methodologies, starting with the Bristol GFM. These assumptions were found to be codified primarily as parameters, which have allowed for delineation of flood maps around the world, through the process which has been termed "from data to hazard". This term also corresponds the general state of all GFMs, which work mostly in ungauged basins, and rather than use local scale data assume all basins and ungauged and derived through remotely sensed. This is unlike local scale inundation studies, where the parameters would only encompass the hydraulic modelling of the domain, the incorporation of remotely sensed data, unconditioned for inundation modelling, requires a series of parameterisations to make the running of a model such as LISFLOOD-FP both possible and realistic. As such, these parameterisations were considered as a part of the overall modelling framework and suitable candidates for sensitivity analysis.

A sensitivity analysis was necessarily the first stage of the thesis, simply because of the number of parameters in the model which were found when the code was reviewed leading to the suspicion that the Bristol GFM is overparameterized. The number of parameters and computational cost associated with running the GFM informed the type of analysis which

would follow. Specifically, the analysis was interested mainly in undertaking a parameter screening with a view to reducing the overall complexity of the model by identifying insensitive parameters (Pianosi et al., 2016). With a smaller pool of parameters, it makes it more like that understanding of the underlying processes of the model becomes tractable (Campolongo et al., 2007). This approach meant the model was setup in a Monte Carlo style experiment, using the Morris method to qualitatively determine which parameters are important (Morris, 1991). Because many of the parameters were based purely on their interactions with the input data, it was deemed important to have several locations to perform the analysis, as the parameters would only be noticeable in the case where that data is present, i.e. vegetation parameters only take an effect within a domain which contains vegetation. Due to the lengthy computation times of the model as well as many internal bugs and errors that needed to be fixed, the time taken up by a Morris method at a single location meant that only 3 were eventually used, although there were many more originally intended. While the conclusions drawn from analysis in chapter 5 were indeed solid enough for the following chapters, an ideal analysis of the parameters at this stage would have involved a greater number of locations, so that correlations regarding the domain and the sensitivity of the parameters could have been more strongly defined.

Nonetheless, the sensitivity analysis successfully reduced the complexity of the GFM, and for 3 locations, we had knowledge of which parameters were sensitive and which were not. While there was an understanding of model variance, it was still to be seen how well the model would perform at a gauged site with data to compare against. For this reason, the Po was considered a good study basin for assessment of the Bristol GFM within a Monte Carlo framework, and a study visit to Bologna with Professor Montanari, who has previously performed detailed studies of the Po (Di Baldassarre & Montanari, 2009; Montanari, 2012), was undertaken. It followed from this that the uncertainty of the parameters could be better understood with a denser sampling scheme, while also being able to assess the model in terms of its skill and realism, by implementing a Monte Carlo methodology similar to the method of Aronica et al., 2002.

Rather than focussing on a single metric for the assessment of the Bristol GFM like previous model evaluation studies, there were instead multiple metrics used to assess the model and permit the identification of parameters from different perspectives (Freer et al., 2004; Savage

et al., 2016). This decision was taken after the initial runs of the model were demonstrably spurious, showing that there was a general lack of knowledge about the domain and how the model should be setup, leading to equifinality and poor parameter identification (Beven & Freer, 2001). This in itself is quite telling about the conditional nature of the modelling undertaken, that there has to be some prior knowledge from the modeller in determining what ranges should be set on parameters and informally set these priors in advance, to reach valid conclusions within a probabilistic framework (Beven et al., 2008). This fact is highly relevant in the context of global flood models, given that their target application is to predominantly ungauged basins. These concerns are indeed what led to the assessment of the model using other metrics not directly related to model performance, namely the ability of the model to reproduce the in-situ river bed of the Po. This was in contrast to the use of flow, area and elevation-based metric used to find more standard indicators of performance. The subsequent peakiness of the posterior distribution of the Manning's coefficient at the expected value is indicative that the model is capable of providing adequate flood estimates without breaking the basic physical realism upon which the model was designed (Clark et al., 2017).

Despite successfully improving our understanding of model behaviour, the studies had thus far failed to find and quantify the greatest source of uncertainty, that of the peak discharge estimation. This evaded us, despite its noted importance being confirmed in the literature (Grimaldi et al., 2019; Zhao et al., 2017). Therefore, it was a surprise to find in the initial screening experiments which were observed, there was no sensitivity to the parameters which were understood to control the peak discharge. As such, this necessarily formed the final research component of the thesis, as it was the greatest source of uncertainty, highlighted by the importance of the proxy parameter of the peak flow multiplier. This parameter was not an initial component of the model however, and was instead a representative for the uncertainty generated by the rFFA, as all the parameters found related to this component of the model were insensitive, giving us the false impression that it was wholly deterministic. As a result, the final chapter began with a complete reassessment of the rFFA method used by the Bristol GFM (Smith et al., 2015), to allow proper quantification of the uncertainty. This found that while the method was grounded in old school methods which have their validity in specific circumstances (Dalrymple, 1960; Meigh et al., 1997), their

implementation at the global scale involved a series of assumptions which led to an incoherent picture of uncertainty.

To mitigate this poor representation of uncertainty, the rFFA was redesigned using an alternative scheme of regionalisation, and the climate classification scheme of Koppen-Geiger was replaced with a continuous indexing scheme of hydroclimatic classification (Knoben et al., 2018), chosen for its ability to represent processes more relevant to flood estimation, such as seasonality (Ahn & Palmer, 2016; Carey et al., 2010; Ouarda et al., 2006; Villarini, 2016). This found that through the use of continuous indices, climates could be clustered in a more credible manner which alongside the new regionalisation scheme effectively parsed and reduced the uncertainty. Another interesting development of this chapter was the emergence of the concept of excess discharge, Q_{EX}. A common limitation of a GFM's reliance on remotely sensed data, is that the bathymetry is often not observable (Alessio Domeneghetti, 2016), and has to be estimated by hydraulic geometry or other morphological relationships (Neal et al., 2015; Wood et al., 2016). This means that in calculation of the peak flows and its propagation over the floodplain, there is some residual uncertainty from the estimation of the bathymetry. Therefore, in finding that the uncertainty of the Q_{EX} is less than the uncertainty of the peak discharge estimates, some of the uncertainty relating to estimation of the bathymetry can be effectively bypassed.

It was noted earlier in the thesis that the Global Flood Model Validation Framework of Hoch & Trigg, 2019 is effective, but had not given sufficient consideration to the myriad uncertainties which beset GFMs. It is hoped that over the course of the results chapters, while mainly focussing on the Bristol GFM, the research conducted with both its successes and failures is able to provide other researchers in global flood modelling with a template of how to better assess their own models, as it is the case that many of the parameterisations and considerations of the modelling framework considered here will be in some way transferable to other GFMs.

8.1. The "SURe" recommendation.

The approach taken throughout the thesis had the aim of better understanding the inner workings of a single GFM, but in the context of a situation where there are multiple GFMs

which need to be better understood. It has also been demonstrated that the lessons learnt about the Bristol GFM are in some cases transferable to other GFMs, as many of the modelling decisions and processes are similar, although this would have to be tested with comparable analyses to confirm. This indicates that the overall approach applied here could be equally beneficial as a methodology for understanding other GFMs. Indeed, it is the case that the sensitivity, uncertainty and refinement of other GFMs will be a significant concern in all future modelling endeavours. It is from this perspective that the concept of "SURe" (Sensitivity, Uncertainty, Refinement) can be recommended to other parts of the global flood modelling community, as a practice to contribute towards the Global Flood Modelling Validation Framework of Hoch & Trigg, 2019.

SURe is a possible template of analysis for global flood modelling. As an iterative process which looks to continuously push for model improvement, the refinement of the models will be constantly addressed as new sources of data are added, different processes are incorporated into the modelling chain, et cetera. For example, in the context of the Bristol GFM, the work done here is far from complete. While quantifying and reducing the uncertainty of the peak discharge is certainly a success, the logical next step would be to reincorporate this back into the modelling setup as a sensitive component itself, rather than being deterministic as it previously was. By performing another sensitivity analysis with this component, it can be seen what the next most erroneous component of the modelling chain is, and this can also become the focus of refinement in its uncertainty and accuracy. This process could continue indefinitely. Recall that there were many simplifications in the initial setup of the model, such as assuming the remotely sensed data is fully deterministic, rather than being subject to significant uncertainty. This decision made the initial study tractable, ensuring that some conclusions could be drawn that weren't mired in equifinality. However, those uncertainties do continue to exist and eventually need to be incorporated, but they will be more manageable and easier to analyse once other components and parameterisations of the model have become well understood. This could therefore be shown in 4 steps:

1. Sensitivity: Calculate the raw sensitivity of the model, selecting the parameters, inputs, or both. This step reduces the complexity of the problem, but makes sure that only the most important of the various components are looked at.

2. Uncertainty: Assess the model with a reduced pool of inputs, in order to address what the effect on model performance is. This will show which model processes are the most important, in the context of model uncertainty which matters.
3. Refinement: After assessing the uncertainty, it should be clear which elements of the model are the most uncertain which when correctly accounted before are the greatest hindrance to model performance. This makes the component therefore the most important one to consider and make improvements to.
4. Repeat: Once an element of the model has been refined, it can be reincorporated back into the framework, so that a new element can be assessed as the most important.

8.2. Missing elements in research

Because of time and computational restraints, an important additional element in further analysis would be to return to the initial screening of the model and perform more sensitivity analyses at various locations. While three analyses were enough to ensure that the most obviously sensitive parameters were identified, there simply aren't enough points sampled to fully understand the interactions of the parameters with the data. A strong hypothesis, confirmed by the 3 study sites shows that as the domains change, so too do the parameter sensitivities. This can only be a result of how the parameters interact with the data to produce an output. As such, there should be an underlying structure which determines parameter sensitivity from the structure of the input data. Given more time and computing resource, this relationship could have been more clearly established.

Similarly, the detailed study at the Po basin, while deeply informative and ultimately leading into the refinement section of the thesis, does not give a definitive account of the associated uncertainties of the model. This is because it was only performed at the Po basin, which while data-rich is nonetheless contextual. Therefore, this method would benefit from being recreated at other locations, although in this case the number of candidate catchments is much smaller, as such a method benefits from the data richness of densely gauged regions, meaning that the studies would most likely be conducted in either North America or Europe.

9. Conclusion

In this chapter, I will briefly re-iterate the main findings of the three results chapters, and bring it together into a synthesis of the overall contribution to the community of global flood modelling. Following this, recommendations for future research are made.

9.1. Main findings

When running the model, there are a small number of fundamental processes which define the modelling chain. In the transition from raw remotely sensed data to inundation modelling, these processes are parameterised. Some parameters were repeatedly important across each modelling domain and they can be summarised as follows:

I. Manning's *n*: This parameter is consistently one of the most important, and this is often also the case in local studies. Fundamentally, the roughness coefficient has a prominent effect on the routing of water along the river reach in all hydrodynamic models. However, in the Bristol GFM, the prediction of channel depth through the use an inverted Manning's equation means that it is also influential to channel bathymetry. Because of its dual function, it is normally the most sensitive parameter.

II. Q_{bf}, bankfull discharge: This parameter is also linked to the hydraulic geometry of the river, as the bankfull return period estimated determines the size of the discharge variable used in the inverted Manning's equation. This value is also significant, because it draws the discharge value from a GEV distribution. This means that the larger Q_{bf} is estimated as the smaller the corresponding value of Q_{EX}, which is what determines the flood hazard. It should be noted that Q_{ex} is dependent on other channel conveyance parameters as well.

III. Reach accumulation threshold: This parameter is responsible for determining the location of boundary conditions throughout the simulated domain, which also means that the model is decomposed into smaller subdomains for individual modelling. The effect of this parameter is that if there are more boundary conditions in the model, there are greater peak flows along the reach, meaning that there is a great overall volume of water being propagated as flood waves throughout the domain. It was a surprise to find this as one of the most important parameters, as it is more of a structural parameter. It was found that the parameter was based on the assumption

of a linear relationship between reach length and upstream accumulation. This means that for the large river basins which we considered, we saw only a discrete change in the outputs, as the accumulation ratio was only significant for a very minor band of the range which we used.

It is possible for global models to be locally relevant. Through Monte Carlo simulations conducted at the Po river, it was seen that the Bristol GFM was capable of producing adequate, behavioural simulations. However, this required some prior knowledge of the site to condition the model setup for effective modelling. For example, the modelling of only main rivers meant knowing to constrain the accumulation threshold parameter, an action which was required given the limited validation data. Furthermore, knowledge of the expected peak flow informed the proxy parameter of peak flow uncertainty. The best results came from the manipulation of the two domain-changing processes within the model, that of modifying the bathymetry and determining the location of the boundary conditions. These are codified by the Manning's n, Q_{bf} and reach accumulation threshold respectively.

The uncertainty of the peak flow is an important consideration for global inundation modelling, and climate is an important predictor variable for peak flow. However, it has been found that Koppen-Geiger has limited descriptive value in hydrological systems when compared with another climate classification system which provides more appropriate hydrologic indicators, such as seasonality and snowfall. While this represents an improvement for a global-scale regionalisation of flood frequency analysis, it has also highlighted the limitations of using only area and average annual rainfall as predictor variables, and incorporation of other catchment descriptors will be necessary to achieve more accurate results.

Overall, this research shows that consideration of the sensitivities and uncertainties of GFMs has genuine benefit for flood hazard assessment, as it can highlight what information should be incorporated for more accurate modelling of the domains.

9.2. Synthesis

The results of this thesis can offer guidelines to other practitioners in the field of global flood modelling on how to assess and improve their own global models. Given the likelihood of

equifinality and overparameterization of the current global models, the most important task is to first reduce model complexity to the most important constituents, and then assess performance under uncertainty. Both of these steps are conducted within a Monte Carlo framework. From these evaluations, one can understand what are the most important aspects of the model to consider, while still being able to challenge their assumptions. The modeller can focus on a specific area of the modelling framework and improve this, understanding the significance that this has for the wider modelling framework. When conducted in an iterative manner, this follows the recommendations of SURe, which can help contribute towards the GFM community's overall goal of cross-model validation and achieving credibility.

The main recommendation for future research rests in understanding the relationships between the global datasets and the related parameters which incorporate them into the modelling framework. This means that the uncertainty of the global datasets has to be understood, so that it can be propagated through to the predicted outputs. By quantifying this uncertainty, it will become possible to provide prior estimates of parameter values, given the contents of the datasets for a modelled location. This will be achieved by producing flood models of many more locations than has been done in this research, requiring work on the model's architecture to facilitate the computational load which could be expected from such an endeavour.

Assessment of the datasets should certainly be a high priority, both in terms of the quantity and quality of data which is available for the estimation of peak flows. The current push in global hydrology to populate databases with gauged stations should be followed very closely and all updates concerning that should be incorporated where possible. Continued assessment of the quality of the gauging data also needs consideration, and what extra information can be leveraged from within. Rather than having access to only climate, area and rainfall, what other variables will help to constrain estimates of peak flow. Alongside this is how to best incorporate uncertainty of the data into the methodology, so that the joint uncertainty of parameters and input data can be understood together, and what interactions are occurring.

It will also be necessary to continue analysis of GFMs at more than a single location, such as the Po. The insights we gathered in chapter 6 are only relevant with respect to the Po river

basin, and the particular characteristics of that domain. As the model is analysed in more domains, then it will become clear how the model constraints vary with respect to different domain characteristics, and will provide a more complete view of how GFMs behave at their native scale.

Acquiring a more holistic view of the behaviour of GFMs should be the ultimate aim of the community. As datasets become more sophisticated and more precise predictions of flooding become available, it will be absolutely crucial for the community to retain its sense of proportion in the meaning of these results. Without a proper appreciation for the provenance of their data the assumptions that their tools make, then it is likely that the comparisons between individual models will never be exhaustively completed, and there will continue to be lingering doubts about the credibility of global flood models.

"To be uncertain is to be uncomfortable, but to be certain is to be ridiculous."

- Socrates

10. References

Abimbola, O. P., Wenninger, J., Venneker, R., & Mittelstet, A. R. (2017). The assessment of water resources in ungauged catchments in Rwanda. *Journal of Hydrology: Regional Studies*, *13*(September 2017), 274–289. https://doi.org/10.1016/j.ejrh.2017.09.001

AdBPo. (2012). *Scenari di rischio residuale*. Retrieved from http://www.adbpo.it/PDGA_Documenti_Piano/PGRA2015/Sezione_A/Relazioni/Materiali_del_Piano/Documento_15_Scenari_di_Rischio_Residuale_Febbraio2012.pdf

Ahilan, S., O'Sullivan, J. J., Bruen, M., Brauders, N., & Healy, D. (2013). Bankfull discharge and recurrence intervals in Irish rivers. *Proceedings of the Institution of Civil Engineers - Water Management*, *166*(7), 381–393. https://doi.org/10.1680/wama.11.00078

Ahn, K. H., & Palmer, R. (2016). Regional flood frequency analysis using spatial proximity and basin characteristics: Quantile regression vs. parameter regression technique. *Journal of Hydrology*, *540*, 515–526. https://doi.org/10.1016/j.jhydrol.2016.06.047

Al-Jazeera. (2019). *Iran suffers $2bn in damages as flood toll continues to rise*. Retrieved from https://www.aljazeera.com/news/2019/04/iran-suffers-2bn-damages-flood-toll-continues-rise-190414085957795.html

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS - global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, *17*(3), 1161–1175. https://doi.org/10.5194/hess-17-1161-2013

Alfieri, L., Burek, P., Feyen, L., & Forzieri, G. (2015). Global warming increases the frequency of river floods in Europe. *Hydrology and Earth System Sciences*, *19*(5), 2247–2260. https://doi.org/10.5194/hess-19-2247-2015

Alfieri, Lorenzo, Salamon, P., Bianchi, A., Neal, J., Bates, P., & Feyen, L. (2014). Advances in pan-European flood hazard mapping. *Hydrological Processes*, *28*(13), 4067–4077. https://doi.org/10.1002/hyp.9947

Ammann, L., Reichert, P., & Fenicia, F. (2018). A framework for likelihood functions of deterministic hydrological models. *Hydrology and Earth System Sciences Discussions*, (August), 1–39. https://doi.org/10.5194/hess-2018-406

Andreadis, K. M., Schumann, G. J. P., & Pavelsky, T. (2013). A simple global river bankfull width and depth database. *Water Resources Research*. https://doi.org/10.1002/wrcr.20440

Apel, H., Thieken, A. H., Merz, B., & Blöschl, G. (2004). Flood risk assessment and associated uncertainty. *Natural Hazards and Earth System Sciences*, *4*, 295–308. Retrieved from https://www.nat-hazards-earth-syst-sci.net/4/295/2004/nhess-4-295-2004.pdf

Arbeláez, A., & Posada, L. (2007). Bankfull discharge in mountain streams in the cauca region of Colombia. *Hydrology Days Conference*, *57*(4). Retrieved from http://hydrologydays.colostate.edu/Papers_2007/Arbelaez_et_al_1_paper.pdf

Arcement, G. J., & Schneider, V. R. (1989). *Guide for Selecting Manning's Roughness Coefficients for Natural Channels and Flood Plains*. https://doi.org/Report No. FHWA-TS-84-204

Aronica, G., Bates, P. D., & Horritt, M. S. (2002). Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrological Processes*, *16*(10), 2001–2016. https://doi.org/10.1002/hyp.398

Bates, P. D., & De Roo, A. P. J. (2000). A simple raster-based model for flood inundation simulation. *Journal of Hydrology*, *236*(1–2), 54–77. https://doi.org/10.1016/S0022-1694(00)00278-X

Bates, Paul D., Horritt, M. S., & Fewtrell, T. J. (2010). A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *Journal of Hydrology*, *387*(1–2), 33–45. https://doi.org/10.1016/j.jhydrol.2010.03.027

Bates, Paul D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., … Krajewski, W. F. (2020). Combined modelling of US fluvial, pluvial and coastal flood hazard under current and future climates. *Water Resources Research*. https://doi.org/10.1029/2020wr028673

Bayliss, A. C., Centre for Ecology and Hydrology, & Institute of Hydrology. (2008). *Flood estimation handbook: Vol. 5: Catchment descriptors*. Wallingford: Centre for Ecology and Hydrology.

Bedri, Z., Bruen, M., Dowley, A., & Masterson, B. (2011). A Three-Dimensional Hydro-Environmental Model of Dublin Bay. *Environmental Modeling and Assessment*, *16*(4), 369–384. https://doi.org/10.1007/s10666-011-9253-7

Benke, K. K., Lowell, K. E., & Hamilton, A. J. (2008). Parameter uncertainty, sensitivity analysis and prediction error in a water-balance hydrological model. *Mathematical and Computer Modelling*, *47*(11–12), 1134–1149. https://doi.org/10.1016/j.mcm.2007.05.017

Berendrecht, W., Weerts, A., Veldhuizen, A., & Kroon, T. (2011). An operational drought forecasting system using coupled models for groundwater, surface water and unsaturated zone. *IAHS-AISH Publication*.

Berghuijs, W. R., Woods, R. A., Hutton, C. J., & Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States. *Geophysical Research Letters*, *43*(9), 4382–4390. https://doi.org/10.1002/2016GL068070

Bernhofen, M. V, Rudari, R., Dottori, F., Sleigh, P. A., Sampson, C. C., Trigg, M. A., … Ward, P. J. (2018). A first collective validation of global fluvial flood models for major floods in Nigeria and Mozambique. *Environmental Research Letters*, *13*(10), 104007. https://doi.org/10.1088/1748-9326/aae014

Beven, K. (2007). Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process. *Hydrology and Earth System Sciences*, *11*(1), 460–467. https://doi.org/10.5194/hess-11-460-2007

Beven, K. J., Smith, P. J., & Freer, J. E. (2008). So just why would a modeller choose to be incoherent? *Journal of Hydrology*, *354*(1–4), 15–32. https://doi.org/10.1016/j.jhydrol.2008.02.007

Beven, Keith. (2000). Uniqueness of place and process representations in hydrological modelling. *Hydrology and Earth System Sciences*, *4*(2), 11. Retrieved from https://www.hydrol-earth-syst-sci.net/4/203/2000/hess-4-203-2000.pdf

Beven, Keith. (2002). Towards a coherent philosophy for modelling the environment. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, *458*(June), 1–20.

Beven, Keith. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1–2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, Keith. (2012). Rainfall-Runoff Modelling: The Primer. In *Wiley-Blackwell*. https://doi.org/10.1002/9781119951001

Beven, Keith, & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305

Beven, Keith, & Binley, A. (2014). GLUE: 20 years on. *Hydrological Processes*, *28*(24), 5897–5918. https://doi.org/10.1002/hyp.10082

Beven, Keith, & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, *249*(1–4), 11–29. https://doi.org/10.1016/S0022-1694(01)00421-8

Beven, Keith, Lamb, R., Leedal, D., & Hunter, N. (2015). Communicating uncertainty in flood inundation mapping: A case study. *International Journal of River Basin Management*, *13*(3), 285–295. https://doi.org/10.1080/15715124.2014.917318

Beven, Keith, Leedal, D., McCarthy, S., Lamb, R., Hunter, N., Keef, C., … Wicks, J. (2011). Framework for assessing uncertainty in fluvial flood risk mapping. In *Flood Risk Management Research Consortium Research Report (available from www. floodrisk. org. uk)*.

Biancamaria, S., Bates, P. D., Boone, A., & Mognard, N. M. (2009). Large-scale coupled hydrologic and hydraulic modelling of the Ob river in Siberia. *Journal of Hydrology*, *379*(1–2), 136–150. https://doi.org/10.1016/j.jhydrol.2009.09.054

Blazkova, S., & Beven, K. (2009). A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resources Research*, *45*(6), 1–12. https://doi.org/10.1029/2007WR006726

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. H. G. (2013). *Runoff Prediction in Ungauged Basins*.

Booker, D. J., & Snelder, T. H. (2012). Comparing methods for estimating flow duration curves at ungauged sites. *Journal of Hydrology*, *434–435*, 78–94. https://doi.org/10.1016/j.jhydrol.2012.02.031

Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, *71*(356), 797–799. https://doi.org/10.1641/b570910

Budyko, M. I. (1974). *Climate and life*. New York: Academic press.

Cameron, D., Beven, K., & Naden, P. (2000). Flood frequency estimation by continuous simulation under climate change (with uncertainty). *Hydrology and Earth System Sciences*, *4*(3), 393–405.

Cameron, D., Beven, K., Tawn, J., & Naden, P. (2000). Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation). *Hydrology and Earth System Sciences*, *4*(1), 23–34.

Cameron, D. S., Beven, K. J., Tawn, J., Blazkova, S., & Naden, P. (1999). Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty). *Journal of Hydrology*, *219*(3–4), 169–187. https://doi.org/10.1016/S0022-1694(99)00057-8

Campolongo, F., Tarantola, S., & Saltelli, A. (1999). Tackling quantitatively large dimensionality problems. *Computer Physics Communications*, *117*(1), 75–85. https://doi.org/10.1016/S0010-4655(98)00165-9

Campolongo, Francesca, Cariboni, J., & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*, *22*(10), 1509–1518. https://doi.org/10.1016/j.envsoft.2006.10.004

Campolongo, Francesca, Saltelli, A., & Cariboni, J. (2011). From screening to quantitative sensitivity analysis. A unified approach. *Computer Physics Communications*, *182*(4), 978–988. https://doi.org/10.1016/j.cpc.2010.12.039

Carey, S. K., Tetzlaff, D., Seibert, J., Soulsby, C., Buttle, J., Laudon, H., … Pomeroy, J. W. (2010). Inter-comparison of hydro-climatic regimes across northern catchments: Synchronicity, resistance and resilience. *Hydrological Processes*, *24*(24), 3591–3602. https://doi.org/10.1002/hyp.7880

Castro, J. M., & Jackson, P. L. (2001). Bankfull discharge recurrence intervals and regional hydraulic geometry relationships: patterns in the Pacific Northwest, USA. *Journal of the American Water Resources Association*, *37*(5), 1249–1262.

Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., … Wagener, T. (2015). Virtual laboratories: New opportunities for collaborative water science. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-19-2101-2015

Chow, V. T., Maidment, D., & Mays, L. W. (1988). *Applied Hydrology*. McGraw Hill.

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., … Peters-Lidard, C. D. (2017). The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. *Hydrology and Earth System Sciences*, *21*(7), 3427–3440. https://doi.org/10.5194/hess-21-3427-2017

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., … Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, *44*(12), 1–14. https://doi.org/10.1029/2007WR006735

Coulthard, T. J., Ramirez, J. A., Barton, N., Rogerson, M., & Brücher, T. (2013). Were rivers flowing across the Sahara during the last interglacial? Implications for human migration through Africa. *PloS One*. https://doi.org/10.1371/journal.pone.0074834

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R., & Smith, P. J. (2015). A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. *Water Resources Research*, *5*(3), 2–2. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Crochemore, L., Isberg, K., Pimentel, R., Pineda, L., Hasan, A., & Arheimer, B. (2019). Lessons learnt from checking the quality of openly accessible river flow data worldwide. *Hydrological Sciences Journal*, *00*(00), 1–13. https://doi.org/10.1080/02626667.2019.1659509

Dalrymple, T. (1960). Flood-Frequency Analyses. Manual of Hydrology Part 3. Flood-flow techniques. *Geological Survey Water-Supply Paper*, *1543-A*, 80. Retrieved from http://pubs.usgs.gov/wsp/1543a/report.pdf

Dankers, R., Arnell, N. W., Clark, D. B., Falloon, P. D., Fekete, B. M., Gosling, S. N., … Wisser, D. (2014). First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. *Proceedings of the National Academy of Sciences*, *111*(9), 3257–3261. https://doi.org/10.1073/pnas.1302078110

Dawdy, D. R., Griffis, V. W., & Gupta, V. K. (2012a). Regional Flood-Frequency Analysis: How We Got Here and Where We Are Going. *Journal of Hydrologic Engineering*, *17*(9), 953–959. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000584

Dawdy, D. R., Griffis, V. W., & Gupta, V. K. (2012b). Regional Flood-Frequency Analysis: How We Got Here and Where We Are Going. *Journal of Hydrologic Engineering*, *17*(9), 953–959. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000584

de Moel, H., Jongman, B., Kreibich, H., Merz, B., Penning-Rowsell, E., & Ward, P. J. (2015). Flood risk assessments at different spatial scales. *Mitigation and Adaptation Strategies for Global Change*, *20*(6), 865–890. https://doi.org/10.1007/s11027-015-9654-z

Di Baldassarre, G., & Montanari, A. (2009). Uncertainty in river discharge observations: A quantitative analysis. *Hydrology and Earth System Sciences*, *13*(6), 913–921. https://doi.org/10.5194/hess-13-913-2009

Di Baldassarre, Giuliano, Castellarin, A., Montanari, A., & Brath, A. (2009). Probability-weighted hazard maps for comparing different flood risk management strategies: a case study. *Natural Hazards*, 479–496. https://doi.org/10.1007/s11069-009-9355-6

Di Baldassarre, Giuliano, Montanari, A., Lins, H., Koutsoyiannis, D., Brandimarte, L., & Blschl, G. (2010). Flood fatalities in Africa: From diagnosis to mitigation. *Geophysical Research Letters*, *37*(22), 2–6. https://doi.org/10.1029/2010GL045467

Di Baldassarre, Giuliano, & Uhlenbrook, S. (2012). Is the current flood of data enough? A treatise on research needs for the improvement of flood modelling. *Hydrological Processes*, *26*(1), 153–158. https://doi.org/10.1002/hyp.8226

Di Miceli, C.M., M.L. Carroll, R.A. Sohlberg, C. Huang, M.C. Hansen, J. R. G. T. (2011). GLCF: MODIS Vegetation Continuous Fields. Retrieved March 13, 2018, from http://glcf.umd.edu/data/vcf/

Dilley, M., Chen, R. S., Deichmann, U., Lerner-Lam, A. L., & Arnold, M. (2005). *Natural Disaster Hotspots: A Global Risk Analysis*. Retrieved from http://hdl.handle.net/10986/7376

Dimitriadis, P., Tegos, A., Oikonomou, A., Pagana, V., Koukouvinos, A., Mamassis, N., … Efstratiadis, A. (2016). Comparative evaluation of 1D and quasi-2D hydraulic models based on benchmark and real-world applications for uncertainty assessment in flood mapping. *Journal of Hydrology*, *534*, 478–492. https://doi.org/10.1016/j.jhydrol.2016.01.020

Domeneghetti, A., Castellarin, A., & Brath, A. (2012). Assessing rating-curve uncertainty and its effects on hydraulic model calibration. *Hydrology and Earth System Sciences*, *16*(4), 1191–1202. https://doi.org/10.5194/hess-16-1191-2012

Domeneghetti, Alessio. (2016). On the use of SRTM and altimetry data for flood modeling in data-sparse regions. *Water Resources Research*, *52*(4), 2901–2918. https://doi.org/10.1002/2015WR017967

Domeneghetti, Alessio, Carisi, F., Castellarin, A., & Brath, A. (2015). Evolution of flood risk over large areas: Quantitative assessment for the Po river. *Journal of Hydrology*, *527*(May), 809–823. https://doi.org/10.1016/j.jhydrol.2015.05.043

Dottori, F., Di Baldassarre, G., & Todini, E. (2013). Detailed data is welcome, but with a pinch of salt: Accuracy, precision, and uncertainty in flood inundation modeling. *Water Resources Research*, *49*(9), 6079–6085. https://doi.org/10.1002/wrcr.20406

Dottori, Francesco, Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F. A., & Feyen, L. (2016). Development and evaluation of a framework for global flood hazard mapping. *Advances in Water Resources*, *94*, 87–102. https://doi.org/10.1016/j.advwatres.2016.05.002

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., … Wood, E. F. (2006). Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, *320*(1–2), 3–17. https://doi.org/10.1016/j.jhydrol.2005.07.031

Eagleson, P. S. (1986). The emergence of global-scale hydrology. *Water Resources Research*, *22*(9 S), 6S-14S. https://doi.org/10.1029/WR022i09Sp0006S

Edwards, P. J., Watson, E. A., & Wood, F. (2019). Toward a Better Understanding of Recurrence Intervals, Bankfull, and Their Importance. *Journal of Contemporary Water Research &*

Education. https://doi.org/10.1111/j.1936-704x.2019.03300.x

Eilander, D., Ikeuchi, H., Yamazaki, D., Couasnon, A., Winsemius, H., & Ward, P. (2018a). Global fluvial flood modelling – a sensitivity analysis. *Geophysical Research Abstracts EGU General Assembly*, *20*, 2018–8397. Retrieved from https://meetingorganizer.copernicus.org/EGU2018/EGU2018-8397.pdf

Eilander, D., Ikeuchi, H., Yamazaki, D., Couasnon, A., Winsemius, H., & Ward, P. (2018b). Global fluvial flood modelling – a sensitivity analysis. *Geophysical Research Abstracts EGU General Assembly*, *20*, 2018–8397. Retrieved from https://meetingorganizer.copernicus.org/EGU2018/EGU2018-8397.pdf

Elvidge, C. D., Tuttle, B. T., Sutton, P. C., Baugh, K. E., Howard, A. T., Milesi, C., … Nemani, R. (2007). Global Distribution and Density of Constructed Impervious Surfaces. *Sensors (Basel, Switzerland)*, *7*(9), 1962–1979. https://doi.org/10.3390/s7091962

EM-DAT. (2018). *Natural Disasters*.

Emerton, R., Zsoter, E., Arnal, L., Cloke, H. L., Muraro, D., Prudhomme, C., … Pappenberger, F. (2018). Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0. *Geoscientific Model Development*, *11*(8), 3327–3346. https://doi.org/10.5194/gmd-11-3327-2018

Encyclopedia Britannica. (2014). Archimedes' Principle. In *The Encyclopedia Britannica*. Retrieved from http://www.britannica.com/EBchecked/topic/32827/Archimedes-principle

Encyclopedia Britannica. (2017). D'Alembert's principle. Retrieved from https://www.britannica.com/science/dAlemberts-principle

Engeland, K., Hisdal, H., & Frigessi, A. (2005). Practical extreme value modelling of hydrological floods and droughts: A case study. *Extremes*, Vol. 7, pp. 5–30. https://doi.org/10.1007/s10687-004-4727-5

Environment Agency. (2010). Flood and Coastal Risk Management Risk Mapping Strategy 2010 – 2015. In *UK Government*.

Euler, L. (1768). *Institutionum calculi integralis*. Petropoli: Impensis Academiae Imperialis Scientarum.

Falconer, R. A. (1986). Water quality simulation study of a natural harbor. *Journal of Waterway, Port, Coastal and Ocean Engineering*. https://doi.org/10.1061/(ASCE)0733-950X(1986)112:1(15)

Farquharson, F. A. K., Meigh, J. R., & Sutcliffe, J. V. (1992). Regional flood frequency analysis in arid and semi-arid areas. *Journal of Hydrology*, *138*(3–4), 487–501. https://doi.org/10.1016/0022-1694(92)90132-F

Fefferman, C. (2006). Existence and smoothness of the Navier-Stokes equations. *The Millennium Prize Problems*.

Fleischmann, A., Paiva, R., & Collischonn, W. (2019). Can regional to continental river hydrodynamic models be locally relevant? A cross-scale comparison. *Journal of Hydrology X*, *3*, 100027. https://doi.org/10.1016/j.hydroa.2019.100027

Franks, S. W., Gineste, P., Beven, K. J., & Merot, P. (1998). On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process. *Water Resources Research*, *34*(4), 787–797. https://doi.org/10.1029/97WR03041

Frasson, R. P. de M., Pavelsky, T. M., Fonstad, M. A., Durand, M. T., Allen, G. H., Schumann, G., …

Yang, X. (2019). Global Relationships Between River Width, Slope, Catchment Area, Meander Wavelength, Sinuosity, and Discharge. *Geophysical Research Letters*, *46*(6), 3252–3262. https://doi.org/10.1029/2019GL082027

Freer, J., Beven, K., & Ambroise, B. (1996). Bayesian esitmation of uncertainty in runoff prediction and the value of data: An application GLUE approach. *Water Resources Research*, *32*(7), 2161–2173.

Freer, J. E., McMillan, H., McDonnell, J. J., & Beven, K. J. (2004). Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2003.12.037

Gale, E. L., & Saunders, M. A. (2013). The 2011 Thailand flood: climate causes and return periods. *Weather*, *68*(9), 233–237. https://doi.org/10.1002/wea.2133

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., … Di, Z. (2014). A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. *Environmental Modelling & Software*, *51*, 269–285. https://doi.org/10.1016/j.envsoft.2013.09.031

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science*, *71*(5), 742–752. https://doi.org/10.1086/425063

Gijsbers, P. J. A., Baayen, J. H., & ter Maat, G. J. (2017). Quick scan tool for water allocation in the Netherlands. *IFIP Advances in Information and Communication Technology*. https://doi.org/10.1007/978-3-319-89935-0_9

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.

Götzinger, J., & Bárdossy, A. (2008). Generic error model for calibration and uncertainty estimation of hydrological models. *Water Resources Research*, *44*(12), 1–18. https://doi.org/10.1029/2007wr006691

GRDC. (2018). Global Runoff Data Base. Retrieved from https://www.bafg.de/GRDC/EN/01_GRDC/13_dtbse/database_node.html

Grimaldi, S., Petroselli, A., Tauro, F., & Porfiri, M. (2012). Time to Concentration: A Paradox of Modern Hydrology. *Hydrological Sciences Journal*, *57*(2), 217–228. https://doi.org/10.1080/02626667.2011.644244

Grimaldi, S., Schumann, G. J. P., Shokri, A., Walker, J. P., & Pauwels, V. R. N. (2019). Challenges, Opportunities, and Pitfalls for Global Coupled Hydrologic-Hydraulic Modeling of Floods. *Water Resources Research*, 5277–5300. https://doi.org/10.1029/2018WR024289

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., & Andréassian, V. (2014). Large-sample hydrology: A need to balance depth with breadth. *Hydrology and Earth System Sciences*, *18*(2), 463–477. https://doi.org/10.5194/hess-18-463-2014

Gupta, Hoshin V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes*, *22*(March), 3802–3813. https://doi.org/10.1002/hyp.6989

Haer, T., Botzen, W. J. W., & Aerts, J. C. J. H. (2019). Advancing disaster policies by integrating dynamic adaptive behaviour in risk assessments using an agent-based modelling approach. *Environmental Research Letters*, *14*(4), 044022. https://doi.org/10.1088/1748-9326/ab0770

Hailegeorgis, T. T., & Alfredsen, K. (2017a). Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. *Journal of*

Hydrology: Regional Studies, *9*, 104–126. https://doi.org/10.1016/j.ejrh.2016.11.004

Hailegeorgis, T. T., & Alfredsen, K. (2017b). Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. *Journal of Hydrology: Regional Studies*, *9*, 104–126. https://doi.org/10.1016/j.ejrh.2016.11.004

Hall, J., & Anderson, M. (2002). Handling uncertainty in extreme or unrepeatable hydrological processes? The need for an alternative paradigm. *Hydrological Processes*, *16*(9), 1867–1870. https://doi.org/10.1002/hyp.5026

Hall, J. W., Tarantola, S., Bates, P. D., & Horritt, M. S. (2005). Distributed Sensitivity Analysis of Flood Inundation Model Calibration. *Journal of Hydraulic Engineering*, *131*(2), 117–126. https://doi.org/10.1061/(ASCE)0733-9429(2005)131:2(117)

Harding, S. G. (1975). *Can Theories Be Refuted? Essays on the Duhem-Quine Thesis*.

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology*, *34*(3), 623–642. https://doi.org/10.1002/joc.3711

Hawker, L., Rougier, J., Neal, J., Bates, P., Archer, L., & Yamazaki, D. (2018). Implications of Simulating Global Digital Elevation Models for Flood Inundation Studies. *Water Resources Research*, *54*(10), 7910–7928. https://doi.org/10.1029/2018WR023279

Hinkel, J., Lincke, D., Vafeidis, A. T., Perrette, M., Nicholls, R. J., Tol, R. S. J., … Levermann, A. (2014). Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1222469111

Hoch, J. M., & Trigg, M. A. (2019). Advancing global flood hazard simulations by improving comparability, benchmarking, and integration of global flood models. *Environmental Research Letters*, *14*(3), 034001. https://doi.org/10.1088/1748-9326/aaf3d3

Hosking, J. R. M., & Wallis, J. R. (1997). L-moments. In *Regional Frequency Analysis: An approa* (pp. 14–43). https://doi.org/10.1017/CBO9780511529443.004

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., … Cudennec, C. (2013). A decade of Predictions in Ungauged Basins (PUB)-a review. *Hydrological Sciences Journal*, *58*(6), 1198–1255. https://doi.org/10.1080/02626667.2013.803183

Huang, P., Li, Z., Chen, J., Li, Q., & Yao, C. (2016). Event-based hydrological modeling for detecting dominant hydrological process and suitable model strategy for semi-arid catchments. *Journal of Hydrology*, *542*, 292–303. https://doi.org/10.1016/j.jhydrol.2016.09.001

Hundecha, Y., Arheimer, B., Donnelly, C., & Pechlivanidis, I. (2016). A regional parameter estimation scheme for a pan-European multi-basin model. *Journal of Hydrology: Regional Studies*, *6*, 90–111. https://doi.org/10.1016/j.ejrh.2016.04.002

Isaak, M. (2002). Flood Stories from Around the World. Retrieved from http://www.talkorigins.org/faqs/flood-myths.html

Jarihani, A. A., Callow, J. N., McVicar, T. R., Van Niel, T. G., & Larsen, J. R. (2015). Satellite-derived Digital Elevation Model (DEM) selection, preparation and correction for hydrodynamic modelling in large, low-gradient and data-sparse catchments. *Journal of Hydrology*, *524*, 489–506. https://doi.org/10.1016/j.jhydrol.2015.02.049

Jingyi, Z., & Hall, M. J. (2004). Regional flood frequency analysis for the Gan-Ming River basin in China. *Journal of Hydrology*, *296*(1–4), 98–117. https://doi.org/10.1016/j.jhydrol.2004.03.018

Jongman, B., Hochrainer-Stigler, S., Feyen, L., Aerts, J. C. J. H., Mechler, R., Botzen, W. J. W., … Ward, P. J. (2014). Increasing stress on disaster-risk finance due to large floods. *Nature Climate Change*, *4*(4), 264–268. https://doi.org/10.1038/nclimate2124

Jongman, B., Winsemius, H. C., Aerts, J. C. J. H., Coughlan de Perez, E., van Aalst, M. K., Kron, W., & Ward, P. J. (2015). Declining vulnerability to river floods and the global benefits of adaptation. *Proceedings of the National Academy of Sciences*, *112*(18), E2271–E2280. https://doi.org/10.1073/pnas.1414439112

Joyce, R. J., Janowiak, J. E., Arkin, P. A., Xie, P., Joyce, R. J., Janowiak, J. E., … Xie, P. (2004). CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *Journal of Hydrometeorology*, *5*(3), 487–503. https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2

Julien, P. Y., & Vensel, C. W. (2005). Review of Sedimentation Issues on the Mississippi River. In *Environmental Engineering*.

Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., & Thielen, J. (2016). Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level. *Environmental Modelling & Software*, *75*, 68–76. https://doi.org/10.1016/j.envsoft.2015.09.009

Kim, K. B., Kwon, H. H., & Han, D. (2018). Exploration of warm-up period in conceptual hydrological modelling. *Journal of Hydrology*, *556*, 194–210. https://doi.org/10.1016/j.jhydrol.2017.11.015

Kjeldsen, T. R., & Jones, D. (2007). Estimation of an index flood using data transfer in the UK. *Hydrological Sciences Journal*. https://doi.org/10.1623/hysj.52.1.86

Knoben, W. J. M., Woods, R. A., & Freer, J. E. (2018). A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data. *Water Resources Research*, *54*(7), 5088–5109. https://doi.org/10.1029/2018WR022913

Koks, E. E., Bočkarjova, M., de Moel, H., & Aerts, J. C. J. H. (2015). Integrated direct and indirect flood risk modeling: Development and sensitivity analysis. *Risk Analysis*, *35*(5), 882–900. https://doi.org/10.1111/risa.12300

Komi, K., Neal, J., Trigg, M. A., & Diekkrüger, B. (2017). Modelling of flood hazard extent in data sparse areas: a case study of the Oti River basin, West Africa. *Journal of Hydrology: Regional Studies*, *10*, 122–132. https://doi.org/10.1016/j.ejrh.2017.03.001

Kumar, R., & Chatterjee, C. (2005). Regional Flood Frequency Analysis Using L-Moments. *Journal of Hydrologic Engineering*, *10*(February), 1–7. https://doi.org/10.1061/(ASCE)1084-0699(2005)10:1(1) CE

Lehner, B., Verdin, K., & Jarvis, A. (2008). New Global Hydrography Derived From Spaceborne Elevation Data. *Eos, Transactions American Geophysical Union*, *89*(10), 93. https://doi.org/10.1029/2008EO100001

Leopold, L. B., & Maddock, T. J. (1953). The Hydraulic Geometry of Stream Channels and Some Physiographic Implications. In *Geological Survey Professional Paper 252*.

Levy, J. K., & Hall, J. (2005). Advances in flood risk management under uncertainty. *Stochastic Environmental Research and Risk Assessment*, *19*(6), 375–377. https://doi.org/10.1007/s00477-005-0005-6

Lin, L., Di, L., Yu, E. G., Kang, L., Shrestha, R., Rahman, M. S., … Hu, L. (2016a). A review of remote sensing in flood assessment. *2016 5th International Conference on Agro-Geoinformatics, Agro-*

Geoinformatics 2016. https://doi.org/10.1109/Agro-Geoinformatics.2016.7577655

Lin, L., Di, L., Yu, E. G., Kang, L., Shrestha, R., Rahman, M. S., … Hu, L. (2016b). A review of remote sensing in flood assessment. *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 1–4. https://doi.org/10.1109/Agro-Geoinformatics.2016.7577655

Liu, Yanli, Freer, J., Beven, K., & Matgen, P. (2009). Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error. *Journal of Hydrology*, *367*(1–2), 93–103. https://doi.org/10.1016/j.jhydrol.2009.01.016

Liu, Yuqiong, & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, *43*(7), 1–18. https://doi.org/10.1029/2006WR005756

M.D. Wilson, & Atkinson, P. . (2003). *Prediction uncertainty in elevation and its effect on flood inundation modelling*.

Manfreda, S., Sole, A., & Di Leo, M. (2011). Detection of Flood-Prone Areas Using Digital Elevation Models. *Journal of Hydrologic Engineering*, *18*(7), 746–759. https://doi.org/10.1061/(ASCE)HE.1943-5584

Mantovan, P., & Todini, E. (2006). Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. *Journal of Hydrology*, *330*(1–2), 368–381. https://doi.org/10.1016/j.jhydrol.2006.04.046

Mateo, C. M. R., Yamazaki, D., Kim, H., Champathong, A., Vaze, J., & Oki, T. (2017). Impacts of spatial resolution and representation of flow connectivity on large-scale simulation of floods. *Hydrology and Earth System Sciences*, *21*(10), 5143–5163. https://doi.org/10.5194/hess-21-5143-2017

Mathworks. (2020). Silhouette plot. Retrieved from Mathworks website: https://uk.mathworks.com/help/stats/silhouette.html#:~:text=A silhouette value measures how,poorly matched to other clusters.

Mcmillan, H. K., & Westerberg, I. K. (2015). Rating curve estimation under epistemic uncertainty. *Hydrological Processes*, *29*(7), 1873–1882. https://doi.org/10.1002/hyp.10419

Meigh, J. (1995). *Regional Flood Estimation Methods for Developing Countries* (pp. 1–143). pp. 1–143.

Meigh, J. R., Farquharson, F. A. K., & Sutcliffe, J. V. (1997). A worldwide comparison of regional flood estimation methods and climate. *Hydrological Sciences Journal*, *42*(2), 225–244. https://doi.org/10.1080/02626669709492022

Merz, R., & Blöschl, G. (2004). Regionalisation of catchment model parameters. *Journal of Hydrology*, *287*(1–4), 95–123. https://doi.org/10.1016/j.jhydrol.2003.09.028

Micklin, P. (2007). The Aral Sea Disaster. *Annual Review of Earth and Planetary Sciences*. https://doi.org/10.1146/annurev.earth.35.031306.140120

Montanari, A. (2012). Hydrology of the Po River: Looking for changing patterns in river discharge. *Hydrology and Earth System Sciences*, *16*(10), 3739–3747. https://doi.org/10.5194/hess-16-3739-2012

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., … Belyaev, V. (2013). "Panta Rhei-Everything Flows": Change in hydrology and society-The IAHS Scientific Decade 2013-2022. *Hydrological Sciences Journal*, *58*(6), 1256–1275. https://doi.org/10.1080/02626667.2013.809088

Montanari, Alberto. (2005). Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, *41*(8), 1–13. https://doi.org/10.1029/2004WR003826

Montanari, Alberto. (2007). What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology. *Hydrological Processes*, *21*(November 2008), 841–845. https://doi.org/10.1002/hyp

Montanari, Alberto. (2011). Uncertainty of Hydrological Predictions. In *Treatise on Water Science* (Vol. 2, pp. 459–478). https://doi.org/10.1016/j.physc.2009.10.149

Montanari, Alberto, Shoemaker, C. A., & Van De Giesen, N. (2009). Introduction to special section on uncertainty assessment in surface and subsurface hydrology: An overview of issues and challenges. *Water Resources Research*, *45*(12), 2005–2008. https://doi.org/10.1029/2009WR008471

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*(2), 161–174. https://doi.org/10.1080/00401706.1991.10484804

Musa, Z. N., Popescu, I., & Mynett, A. (2015). A review of applications of satellite SAR, optical, altimetry and DEM data for surface water modelling, mapping and parameter estimation. *Hydrology and Earth System Sciences*, *19*(9), 3755–3769. https://doi.org/10.5194/hess-19-3755-2015

Neal, J. C., Odoni, N. A., Trigg, M. A., Freer, J. E., Garcia-Pintado, J., Mason, D. C., … Bates, P. D. (2015). Efficient incorporation of channel cross-section geometry uncertainty into regional and global scale flood inundation models. *Journal of Hydrology*, *529*, 169–183. https://doi.org/10.1016/j.jhydrol.2015.07.026

Neal, J., Schumann, G., & Bates, P. (2012). A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. *Water Resources Research*, *48*(11), 1–16. https://doi.org/10.1029/2012WR012514

Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., & Weijs, S. V. (2016). A philosophical basis for hydrological uncertainty. *Hydrological Sciences Journal*, *61*(9), 1666–1678. https://doi.org/10.1080/02626667.2016.1183009

NERC. (1975). *Flood Studies Report*. NERC.

Newton, I. (1687). *Philosophiae Naturalis Principia Mathematica*.

Nguyen, T. G., & de Kok, J. L. (2007). Systematic testing of an integrated systems model for coastal zone management using sensitivity and uncertainty analyses. *Environmental Modelling and Software*, *22*(11), 1572–1587. https://doi.org/10.1016/j.envsoft.2006.08.008

O'Loughlin, F. E., Paiva, R. C. D., Durand, M., Alsdorf, D. E., & Bates, P. D. (2016). A multi-sensor approach towards a global vegetation corrected SRTM DEM product. *Remote Sensing of Environment*, *182*, 49–59. https://doi.org/10.1016/j.rse.2016.04.018

Ocio, D., Le Vine, N., Westerberg, I., Pappenberger, F., & Buytaert, W. (2017). The role of rating curve uncertainty in real-time flood forecasting. *Water Resources Research*, *53*(5), 4197–4213. https://doi.org/10.1002/2016WR020225

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, *263*(5147), 641–646. https://doi.org/10.1126/science.263.5147.641

Ouarda, T. B. M. J., Cunderlik, J. M., St-Hilaire, A., Barbet, M., Bruneau, P., & Bobée, B. (2006). Data-

based comparison of seasonality-based regional flood frequency methods. *Journal of Hydrology*, *330*(1–2), 329–339. https://doi.org/10.1016/j.jhydrol.2006.03.023

Pappenberger, F., Dutra, E., Wetterhall, F., & Cloke, H. L. (2012). Deriving global flood hazard maps of fluvial floods through a physical model cascade. *Hydrology and Earth System Sciences*, *16*(11), 4143–4156. https://doi.org/10.5194/hess-16-4143-2012

Pappenberger, Florian, Beven, K. J., Ratto, M., & Matgen, P. (2008). Multi-method global sensitivity analysis of flood inundation models. *Advances in Water Resources*, *31*(1), 1–14. https://doi.org/10.1016/j.advwatres.2007.04.009

Parkes, B., & Demeritt, D. (2016). Defining the hundred year flood: A Bayesian approach for using historic data to reduce uncertainty in flood frequency estimates. *Journal of Hydrology*, *540*, 1189–1208. https://doi.org/10.1016/j.jhydrol.2016.07.025

Pascal, B. (1663). *Traitez de l'Equilibre des Liqueurs*.

Peel, M. C., Finlayson, B. L., & McMahon, T. a. (2006). Updated world map of the K ̈oppen-Geiger climate classification. *Meteorologische Zeitschrift*, *15*, 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Peel, Murray C., McMahon, T. A., & Finlayson, B. L. (2004). Continental differences in the variability of annual runoff-update and reassessment. *Journal of Hydrology*, *295*(1–4), 185–197. https://doi.org/10.1016/j.jhydrol.2004.03.004

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

Petit, F., & Pauquet, A. (1997). Bankfull discharge recurrence interval in gravel-bed rivers. *Earth Surface Processes and Landforms*, *22*(7), 685–693. https://doi.org/10.1002/(SICI)1096-9837(199707)22:7<685::AID-ESP744>3.0.CO;2-J

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling and Software*, *79*, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Pianosi, F., Sarrazin, F., & Wagener, T. (2015). A Matlab toolbox for Global Sensitivity Analysis. *Environmental Modelling and Software*, *70*, 80–85. https://doi.org/10.1016/j.envsoft.2015.04.009

Ponnamperuma, F. N. (1984). Effects of Flooding on Soils. In *Flooding and Plant Growth*. https://doi.org/10.1016/b978-0-12-424120-6.50007-9

Popper, K. R. (1963). Science as Falsification. In *Conjectures and Refutations* (pp. 1–6). https://doi.org/10.1108/00330330310484369

Razavi, S., & Gupta, H. (2016). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. *Water Resource Reasearch*, *52*, 423–439. https://doi.org/10.1002/2015WR017559.A

Razavi, S., & Gupta, H. V. (2015). What do we mean by sensitivity analysis? The need for comprehensive characterization of '"global"' sensitivity in Earth and Environmental systems models. *Water Resources Research*, *5*(3), 2–2. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Renard, B., Kochanek, K., Lang, M., Garavaglia, F., Paquet, E., Neppel, L., … Auffray, A. (2013). Data-

based comparison of frequency analysis methods: A general framework. *Water Resources Research*, *49*(2), 825–843. https://doi.org/10.1002/wrcr.20087

Renard, Benjamin, Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*(5), 1–22. https://doi.org/10.1029/2009WR008328

Robson, A. J., Houghton-Carr, H. A., Faulkner, D. S., Jakob, D., Marshall, D. C. W., & Reed, D. W. (1999). The Flood Estimation Handbook methods. *Proceedings of the 1999 34th MAFF Conference of River and Coastal Engineers*.

Rodríguez, E., Morris, C. S., & Belz, J. E. (2006). A Global Assessment of the SRTM Performance. *Photogrammetric Engineering & Remote Sensing*, *72*(3), 249–260. https://doi.org/10.14358/PERS.72.3.249

Romanowicz, R., Beven, K. J., & Tawn, J. (1996). Bayesian calibration of flood inundation models. In *Floodplain Processes*.

Rudari, R., Silvestro, F., Campo, L., Rebora, N., Boni, G., & Herold, C. (2015). *Improvement of the global flood model for the GAR 2015*.

Saint-Venant, J. C. B. de. (1871). Théorie du mouvement non permanent des eaux, avec application aux crues des rivières et a l'introduction de marées dans leurs litsNo Title. *Comptes Rendus de l'Académie Des Sciences*, *73*.

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins-Part 2: Flood and low flow studies. *Hydrology and Earth System Sciences*, *17*(7), 2637–2652. https://doi.org/10.5194/hess-17-2637-2013

Saltelli, A., & Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling and Software*, *25*(12), 1508–1517. https://doi.org/10.1016/j.envsoft.2010.04.012

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, *181*(2), 259–270. https://doi.org/10.1016/j.cpc.2009.09.018

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., … Tarantola, S. (2007). *Global Sensitivity Analysis. The Primer*. https://doi.org/10.1002/9780470725184

Samela, C., Troy, T. J., & Manfreda, S. (2017). Geomorphic classifiers for flood-prone areas delineation for data-scarce environments. *Advances in Water Resources*, *102*, 13–28. https://doi.org/10.1016/j.advwatres.2017.01.007

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., & E., F. J. (2015). A high-resolution global flood hazard model. *Water Resources Research*, 7358–7381. https://doi.org/10.1002/2015WR016954.Received

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., & Trigg, M. A. (2016, January 5). Perspectives on Open Access High Resolution Digital Elevation Models to Produce Global Flood Hazard Layers. *Frontiers in Earth Science*, *3*. https://doi.org/10.3389/feart.2015.00085

Sandesara, U., & Wooten, T. (1986). *No one had a tongue to speak: the untold story of one of history's deadliest floods*.

Sanyal, J., Carbonneau, P., & Densmore, A. L. (2014). Low-cost inundation modelling at the reach scale with sparse data in the Lower Damodar River basin, India. *Hydrological Sciences Journal*,

59(12), 2086–2102. https://doi.org/10.1080/02626667.2014.884718

Sanyal, J., & Lu, X. X. (2004). Application of Remote Sensing in Flood Management with Special Reference to Monsoon Asia: A Review. *Natural Hazards*, *33*(2), 283–301. https://doi.org/10.1023/B:NHAZ.0000037035.65105.95

Sarrazin, F., Pianosi, F., & Wagener, T. (2016). Global Sensitivity Analysis of environmental models: Convergence and validation. *Environmental Modelling and Software*, *79*, 135–152. https://doi.org/10.1016/j.envsoft.2016.02.005

Savage, J., Pianosi, F., Bates, P., Freer, J., & Wagener, T. (2016). Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model. *Water Resources Research*, *52*(11), 9146–9163. https://doi.org/10.1002/2015WR018198.Received

Savage, J. T. S., Bates, P., Freer, J., Neal, J., & Aronica, G. (2016). When does spatial resolution become spurious in probabilistic flood inundation predictions? *Hydrological Processes*, *30*(13), 2014–2032. https://doi.org/10.1002/hyp.10749

Schoups, G., Van De Giesen, N. C., & Savenije, H. H. G. (2008). Model complexity control for hydrologic prediction. *Water Resources Research*, *44*(1), 1–14. https://doi.org/10.1029/2008WR006836

Schumann, G. J. P., Andreadis, K. M., & Bates, P. D. (2014). Downscaling coarse grid hydrodynamic model simulations over large domains. *Journal of Hydrology*, *508*, 289–298. https://doi.org/10.1016/j.jhydrol.2013.08.051

Schumann, G. J. P., Bates, P. D., Apel, H., & Aronica, G. (2018). *Global Flood Hazard: Applications in Modeling, Mapping, and Forecasting*. Wiley.

Schumann, G. J. P., Bates, P. D., Neal, J. C., & Andreadis, K. M. (2014, March 12). Technology: Fight floods on a global scale. *Nature*, Vol. 507, p. 169. https://doi.org/10.1038/507169e

Schumann, G. J. P., Stampoulis, D., Smith, A. M., Sampson, C. C., Andreadis, K. M., Neal, J. C., & Bates, P. D. (2016). Rethinking flood hazard at the global scale. *Geophysical Research Letters*, *43*(19), 10,249-10,256. https://doi.org/10.1002/2016GL070260

Schwanghart, W., & Kuhn, N. J. (2010). TopoToolbox: A set of Matlab functions for topographic analysis. *Environmental Modelling and Software*. https://doi.org/10.1016/j.envsoft.2009.12.002

Scussolini, P., Aerts, J. C. J. H., Jongman, B., Bouwer, L. M., Winsemius, H. C., de Moel, H., & Ward, P. J. (2016). FLOPROS: an evolving global database of flood protection standards. *Natural Hazards and Earth System Sciences*, *16*(5), 1049–1061. https://doi.org/10.5194/nhess-16-1049-2016

Sharifi, L., & Bokaie, S. (2019). Priorities in prevention and control of flood hazards in Iran 2019 massive flood. *Iranian Journal of Microbiology*, *11*(2), 80–84. https://doi.org/10.18502/ijm.v11i2.1065

Shastry, A., & Durand, M. (2019). Utilizing flood inundation observations to obtain floodplain topography in data-scarce regions. *Frontiers in Earth Science*, *6*(January), 1–10. https://doi.org/10.3389/feart.2018.00243

Sikorska, A., Viviroli, D., & Seibert, J. (2015). Flood-type classification in mountainous catchments using crisp and fuzzy decision trees. *Water Resources Research*, *51*, 6725–6738. https://doi.org/10.1002/2015WR017618.Received

Simard, M., Pinto, N., Fisher, J. B., & Baccini, A. (2011). Mapping forest canopy height globally with

spaceborne lidar. *Journal of Geophysical Research*, *116*(G4), G04021. https://doi.org/10.1029/2011JG001708

Singh, R., Wagener, T., Van Werkhoven, K., Mann, M. E., & Crane, R. (2011). A trading-space-for-time approach to probabilistic continuous streamflow predictions in a changing climate-accounting for changing watershed behavior. *Hydrology and Earth System Sciences*, *15*(11), 3591–3603. https://doi.org/10.5194/hess-15-3591-2011

Siqueira, V. A., Paiva, R. C. D., Fleischmann, A. S., Fan, F. M., Ruhoff, A. L., Pontes, P. R. M., … Collischonn, W. (2018). Toward continental hydrologic-hydrodynamic modeling in South America. *Hydrology and Earth System Sciences*, *22*(9), 4815–4842. https://doi.org/10.5194/hess-22-4815-2018

Smith, A., Sampson, C., & Bates, P. (2015). Regional flood frequency analysis at the global scale. *Water Resour.Res.*, *51*, 539–553. https://doi.org/10.1002/2014WR015814.Received

Smith, Andrew, Bates, P. D., Wing, O., Sampson, C., Quinn, N., & Neal, J. (2019). New estimates of flood exposure in developing countries using high-resolution population data. *Nature Communications*, *10*(1), 1–7. https://doi.org/10.1038/s41467-019-09282-y

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., & Xu, C. (2015). Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. *Journal of Hydrology*, *523*, 739–757. https://doi.org/10.1016/j.jhydrol.2015.02.013

Sood, A., & Smakhtin, V. (2015). Global hydrological models: a review. *Hydrological Sciences Journal*, *60*(4), 549–565. https://doi.org/10.1080/02626667.2014.950580

Stedinger, J. R., Vogel, R. M., Lee, S. U., & Batchelder, R. (2008). Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resources Research*, *44*(12), 1–17. https://doi.org/10.1029/2008WR006822

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K. L. (2018). A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, *56*(1), 79–107. https://doi.org/10.1002/2017RG000574

Tanoue, M., Hirabayashi, Y., & Ikeuchi, H. (2016). Global-scale river flood vulnerability in the last 50 years. *Scientific Reports*, *6*(1), 36021. https://doi.org/10.1038/srep36021

Tarasova, L., Merz, R., Kiss, A., Basso, S., Blöschl, G., Merz, B., … Wietzke, L. (2019). Causative classification of river flood events. *Wiley Interdisciplinary Reviews: Water*, (January), e1353. https://doi.org/10.1002/wat2.1353

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling & Software*, *90*, 201–216. https://doi.org/10.1016/j.envsoft.2017.01.006

Teng, J., Vaze, J., Dutta, D., & Marvanek, S. (2015). Rapid Inundation Modelling in Large Floodplains Using LiDAR DEM. *Water Resources Management*, *29*(8), 2619–2636. https://doi.org/10.1007/s11269-015-0960-8

Toombes, L., & Chanson, H. (2011). Numerical Limitations of Hydraulic Models. *Journal of Hydrology*.

Tourian, M. J., Elmi, O., Mohammadnejad, A., & Sneeuw, N. (2017). Estimating river depth from SWOT-type observables obtained by satellite altimetry and imagery. *Water (Switzerland)*, *9*(10). https://doi.org/10.3390/w9100753

Trigg, M. A., Birch, C. E., Neal, J. C., Bates, P. D., Smith, A., Sampson, C. C., … Fewtrell, T. J. (2016). The credibility challenge for global fluvial flood risk analysis. *Environmental Research Letters*, *11*(9),

1–10. https://doi.org/10.1088/1748-9326/11/9/094014

Trigg, Mark A., Bates, P. D., Wilson, M. D., Schumann, G., & Baugh, C. (2012). Floodplain channel morphology and networks of the middle Amazon River. *Water Resources Research*, *48*(10), 1–17. https://doi.org/10.1029/2012WR011888

UNISDR. (2015). *The Pocket GAR 2015 Making Development Sustainable: The Future of Disaster Risk Management*. Retrieved from https://www.preventionweb.net/english/hyogo/gar/2015/en/gar-pdf/GAR15_Pocket_EN.pdf

USACE. (2010). HEC-RAS River Analysis System. *User's Manual, Version 4.1*. https://doi.org/CPD-68

Valiani, A., Caleffi, V., & Zanni, A. (2002). Case Study: Malpasset Dam-Break Simulation using a Two-Dimensional Finite Volume Method. *Journal of Hydraulic Engineering*, *125*(5), 460–472. https://doi.org/10.1061/~ASCE!0733-9429~2002!128:5~460!

van Werkhoven, K., Wagener, T., Reed, P., & Tang, Y. (2008). Characterization of watershed model behavior across a hydroclimatic gradient. *Water Resources Research*, *44*(1), 1–16. https://doi.org/10.1029/2007WR006271

Vemuri, V., Dracup, J. A., Erdmann, R. C., & Vemubi, N. (1969). Sensitivity Analysis Method of System Identification and Its Potential in Hydrologic Research. *Water Resources Research*, *5*(2). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/WR005i002p00341

Viglione, A., Merz, R., Luis Salinas, J., & Blöschl, unter. (2013). Flood frequency hydrology: 3. A Bayesian analysis. *Water Resour. Res*, *49*. https://doi.org/10.1029/2011WR010782

Villarini, G. (2016). On the seasonality of flooding across the continental United States. *Advances in Water Resources*, *87*, 80–91. https://doi.org/10.1016/j.advwatres.2015.11.009

Vorogushyn, S., Bates, P. D., de Bruijn, K., Castellarin, A., Kreibich, H., Priest, S., … Merz, B. (2018). Evolutionary leap in large-scale flood risk assessment needed. *Wiley Interdisciplinary Reviews: Water*, *5*(2), e1266. https://doi.org/10.1002/wat2.1266

Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V., & Robinson, B. A. (2009). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, *23*(7), 1011–1026. https://doi.org/10.1007/s00477-008-0274-y

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., & Gupta, H. V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrological Processes*, *17*(2), 455–476. https://doi.org/10.1002/hyp.1135

Wagener, Thorsten, & Gupta, H. V. (2005). Model identification for hydrological forecasting under uncertainty. *Stochastic Environmental Research and Risk Assessment*, *19*(6), 378–387. https://doi.org/10.1007/s00477-005-0006-5

Wagener, Thorsten, Sivapalan, M., Troch, P., & Woods, R. (2007). Catchment Classification and Hydrologic Similarity. *Geography Compass*, *1*(4), 901–931. https://doi.org/10.1111/j.1749-8198.2007.00039.x

Wainwright, H. M., Finsterle, S., Jung, Y., Zhou, Q., & Birkholzer, J. T. (2014). Making sense of global sensitivity analyses. *Computers and Geosciences*, *65*, 94–94. https://doi.org/10.1016/j.cageo.2013.06.006

Ward, P. J., Jongman, B., Salamon, P., Simpson, A., Bates, P., De Groeve, T., … Winsemius, H. C. (2015). Usefulness and limitations of global flood risk models. *Nature Climate Change*, *5*(8), 712–715. https://doi.org/10.1038/nclimate2742

Ward, P. J., Jongman, B., Weiland, F. S., Bouwman, A., Van Beek, R., Bierkens, M. F. P., … Winsemius, H. C. (2013). Assessing flood risk at the global scale: Model setup, results, and sensitivity. *Environmental Research Letters*, *8*(4). https://doi.org/10.1088/1748-9326/8/4/044019

Warren, I. R., & Bach, H. K. (1992). MIKE 21: a modelling system for estuaries, coastal waters and seas. *Environmental Software*. https://doi.org/10.1016/0266-9838(92)90006-P

WBM, B. (2001). TUFLOW User Manual. *6th Conference on Hydraulics in Civil Engineering: The State of Hydraulics*.

Weichel, T., Pappenberger, F., & Schulz, K. (2007). Sensitivity and uncertainty in flood inundation modelling – concept of an analysis framework. *Advances in Geosciences*, *11*, 31–36. https://doi.org/10.5194/adgeo-11-31-2007

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., & Freer, J. (2016). Uncertainty in hydrological signatures for gauged and ungauged catchments. *Water Resources Research*, *52*(3), 1847–1865. https://doi.org/10.1002/2015WR017635

Wharton, G., & Tomlinson, J. J. (1999). Flood discharge estimation from river channel dimensions: results of applications in Java, Burundi, Ghana and Tanzania / Estimation d'un débit de crue à partir de la géométrie du chenal d'écoulement: applications et résultats à Java, au Burundi, au Ghana . *Hydrological Sciences Journal*, *44*(1), 97. https://doi.org/10.1080/02626669909492205

WHS. (2014). WinFAP. https://doi.org/https://www.hydrosolutions.co.uk/software/winfap-4/

Williams, G. P. (1978). Bank-full discharge of rivers. *Water Resources Research*, *14*(6), 1141–1154. https://doi.org/10.1029/WR014i006p01141

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., … Xavier, P. K. (2018). The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. *Journal of Advances in Modeling Earth Systems*. https://doi.org/10.1002/2017MS001115

Wing, O. E. J., Bates, P. D., Neal, J. C., Sampson, C. C., Smith, A. M., Quinn, N., … Krajewski, W. F. (2019). A new automated method for improved flood defense representation in large-scale hydraulic models. *Water Resources Research*, 0–2. https://doi.org/10.1029/2019WR025957

Wing, O. E. J., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., & Erickson, T. A. (2017). Validation of a 30 m resolution flood hazard model of the conterminous United States. *Water Resour. Res*. https://doi.org/10.1002/2017WR020917

Wing, O. E. J., Bates, P. D., Smith, A. M., Sampson, C. C., Johnson, K. A., Fargione, J., & Morefield, P. (2018). Estimates of present and future flood risk in the conterminous United States. *Environmental Research Letters*, *13*(3), 034023. https://doi.org/10.1088/1748-9326/aaac65

Winsemius, H. C., Van Beek, L. P. H., Jongman, B., Ward, P. J., & Bouwman, A. (2013). A framework for global river flood risk assessments. *Hydrology and Earth System Sciences*, *17*(5), 1871–1892. https://doi.org/10.5194/hess-17-1871-2013

Winsemius, Hessel C., Aerts, J. C. J. H., van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., … Ward, P. J. (2016). Global drivers of future river flood risk. *Nature Climate Change*, *6*(4), 381–385. https://doi.org/10.1038/nclimate2893

Winsemius, Hessel C., Ward, P. J., & Luo, T. (2015). Aqueduct Global Flood Risk Maps.

Winter, T. C. (2001). The Concept of Hydrologic Landscapes. *Journal of the American Water Resources Association*, *37*(2), 335–349. https://doi.org/10.1111/j.1752-1688.2001.tb00973.x

Wood, M., Hostache, R., Neal, J., Wagener, T., Giustarini, L., Chini, M., … Bates, P. (2016). Calibration

of channel depth and friction parameters in the LISFLOOD-FP hydraulic model using medium-resolution SAR data and identifiability techniques. *Hydrology and Earth System Sciences*, *20*(12), 4983–4997. https://doi.org/10.5194/hess-20-4983-2016

World Bank. (2012). *Thai Flood 2011: Rapid Assessment for Resilient Recovery and Reconstruction Planning*. Retrieved from http://documents.worldbank.org/curated/en/677841468335414861/pdf/698220WP0v10P106011020120Box370022B.pdf

Yadav, M., Wagener, T., & Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, *30*(8), 1756–1774. https://doi.org/10.1016/j.advwatres.2007.01.005

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., … Bates, P. D. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, *44*(11), 5844–5853. https://doi.org/10.1002/2017GL072874

Yamazaki, D., Kanae, S., Kim, H., & Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*, *47*(4), 1–21. https://doi.org/10.1029/2010WR009726

Yamazaki, D., O'Loughlin, F., Trigg, M. A., Miller, Z. F., Pavelsky, T. M., & Bates, P. D. (2014). Development of the Global Width Database for Large Rivers. *Water Resources Research*, *50*(4), 3467–3480. https://doi.org/10.1002/2013WR014664

Yamazaki, D., Sato, T., Kanae, S., Hirabayashi, Y., & Bates, P. D. (2014). Regional flood dynamics in a bifurcating mega delta simulated in a global river model. *Geophysical Prospecting*, *41*(8), 2928–2933. https://doi.org/10.1002/2014GL061184.Received

Yan, K., Di Baldassarre, G., Solomatine, D. P., & Schumann, G. J.-P. (2015). A review of low-cost space-borne data for flood modelling: topography, flood extent and water level. *Hydrological Processes*, *29*(15), 3368–3387. https://doi.org/10.1002/hyp.10449

Yan, K., Tarpanelli, A., Balint, G., Moramarco, T., & Baldassarre, G. Di. (2015). Exploring the Potential of SRTM Topography and Radar Altimetry to Support Flood Propagation Modeling: Danube Case Study. *Journal of Hydrologic Engineering*, *20*(2), 04014048. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001018

Zaman, M. A., Rahman, A., & Haddad, K. (2012). Regional flood frequency analysis in arid regions: A case study for Australia. *Journal of Hydrology*, *475*, 74–83. https://doi.org/10.1016/j.jhydrol.2012.08.054

Zhao, F., Veldkamp, T. I. E., Frieler, K., Schewe, J., Ostberg, S., Willner, S., … Yamazaki, D. (2017). The critical role of the routing scheme in simulating peak river discharge in global hydrological models. *Environmental Research Letters*, *12*(7), 075003. https://doi.org/10.1088/1748-9326/aa7250